

(TM)

1	134	1.6	509	1	P91060	Predicted amino acid	3.44e-00
2	125	1.5	402	26	W58855	Full length sequence	1.45e-00
3	125	1.5	402	10	R30238	Human Op-2.	1.45e-00
4	125	1.5	402	10	R54937	Osteogenic protein hO	1.45e-00
5	125	1.5	402	11	R60578	Osteogenic protein hO	1.45e-00
6	125	1.5	402	9	R57973	Human Op-2.	1.45e-00
7	125	1.5	402	5	R27291	Human osteogenic prot	1.45e-00
8	125	1.5	402	9	R45732	Human osteogenic pro-	1.45e-00
9	125	1.5	402	22	W15366	Human osteogenic prot	1.45e-00
10	125	1.5	402	8	R4759	Human hippocampal os	1.45e-00
11	125	1.5	402	6	R30410	Human osteogenic prot	1.45e-00
12	125	1.5	402	19	W00238	Human Op-2-PP.	1.45e-00
13	125	1.5	402	9	R47292	Human osteogenic prot	1.45e-00
14	125	1.5	402	12	R51647	hOP2.	1.45e-00
15	125	1.5	402	12	R52000	hOP2-PP prepro form h	1.45e-00
16	125	1.5	402	9	R50200	Human Op-2.	1.45e-00
17	125	1.5	402	9	R47252	hOP2.	1.45e-00
18	124	1.5	402	16	R85760	hOP-2.	1.45e-00
19	123	1.5	439	5	R3907	Human osteogenic prot	1.98e-00
					R28150	Sugar beet chitinase	1.98e-00

Db 96 tpppappedrpgagnasrdgrpsgggrprprprskappkxkw 140
 QY 274 SPARPAEATSLGALSGTRHSVGVGRQHAGPSTSRP-PRPW 317

RESULT 2
 ID W36855 standard; Protein; 402 AA.
 AC W36855;
 DT 11-OCT-1994 (first entry)
 DE Full length sequence of human osteogenic protein 2 (hOP-2).
 KW Human osteogenic protein; OP; OP-2; morphogen; morphogenic protein;
 KW embryogenesis; organ maintenance; tissue-specific morphogenesis;
 KW arthritis; emphysema; osteoporosis; cirrhosis.
 OS Homo sapiens.
 FH Key
 FT Region
 FT 18..263
 /note= "pro region which is cleaved to yield the mature morphogenically active protein"
 Protein 264..402
 Region /note= "mature protein"
 301..402
 /note= "conserved 7 Cys skeleton"

US5650276-A.
 22-JUL-1997. 278729.
 PD 20-JUL-1994; US-938021.
 PR 28-AUG-1992; US-752764.
 PR 30-AUG-1991; US-752861.
 PR 30-AUG-1991; US-752861.
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;
 PI Pang RH, Rueger DC, Smart JE;
 PI WPI: 97-384665/35.
 DR N-PSDB; T97881.
 DR Screening for compounds which modulate morphogen expression - by
 PT incubating in the presence of epithelial cells which contain a
 PT cellular gene for morphogenic protein expression
 PS Disclosure; Columns 61-64; 49pp; English.
 CC The present sequence represents a human osteogenic protein-2 (hOP-2).
 CC OP-2 proteins are a group of morphogenically active proteins. Morphogens
 CC are inactive when reduced, but are active as oxidised homodimers and when
 CC oxidised with other morphogens (e.g. W36853-02). Comparison of the amino
 CC acid sequences of these morphogens has identified a consensus 6-7
 CC cysteine motif at the C-terminal. Morphogenic proteins such as OP-2 play
 CC an important role, not only in embryogenesis, but also in tissue and
 CC organ maintenance and repair in mammals. They induce a developmental
 CC cascade of tissue-specific morphogenesis in a mammal. A novel method is
 CC described for screening a candidate compound for the ability to modulate
 CC expression of a cellular gene encoding a naturally occurring morphogenic
 CC protein. The candidate compound is incubated with epithelial cells which
 CC express the cellular gene, and after a period of time the epithelial
 CC cells are assayed for the presence of or the amount of the protein
 CC expressed by the cellular gene. A change in the level of the
 CC morphogenic protein relative to the level in the epithelial cells in the
 CC absence of the candidate compound is indicative of the ability of the
 CC compound to modulate expression of the cellular gene. The method can be
 CC used to identify compounds which can increase or decrease morphogen
 CC production or levels. Such compounds can be used in the treatment of,
 CC e.g. arthritis, emphysema, osteoporosis, kidney disease, lung diseases,
 CC cardiomyopathy, and cirrhosis of the liver.
 SQ Sequence 402 AA;

Query Match 1.5%; Score 125; DB 26; Length 402;
 Best Local Similarity 39.7%; Pred. No. 1.45e+00;
 Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggglrppp-gc-pqrrlg-arerrdvqrellavlgpr-prprappaas 67
 QY 172 GPPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRGGASR 230
 Db 68 rlp 70
 QY 231 SLP 233

RESULT 3
 ID R50238 standard; Protein; 402 AA.
 AC R50238;
 DT 11-OCT-1994 (first entry)
 DE Human OP-2.
 KW OP-1; OP-2; CBMP2; Vgl(fx); Vgr(fx); DPP(fx);
 KW GDF-1(fx); 60A(fx); BMP3(fx); BMP5(fx); BMP6(fx);
 KW osteogenic protein; morphogen; morphogenic protein;
 KW liver; regeneration; injury; cancer; integration;
 KW transplant; gene therapy; hepatic tissue.
 OS Homo sapiens.
 PN WO9406449-A.
 PD 31-MAR-1994.
 PF 16-SEP-1993; U08808.
 PR 16-SEP-1992; US-946238.
 PR 04-MAR-1993; US-029335.
 PR 31-MAR-1993; US-040510.
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;
 PI Pang RH, Rueger DC, Smart JE;
 PI WPI: 94-118148/14.
 DR N-PSDB; Q45164.
 DR Use of morphogen(s) to induce liver regeneration - for repair of
 PT injury, treatment of cancer etc., also to improved integration of
 PT transplant tissue, in gene therapy etc..
 PS Claim 42-42; Page 134-136; 176pp; English.
 CC Morphogens comprising an amino acid sequence sharing at least
 CC 70% homology with OP-1, OP-2, CBMP2, Vgl(fx), Vgr(fx), DPP(fx),
 CC GDF-1(fx), 60A(fx) are useful for maintaining liver function in
 CC a mammal, including means for regenerating lost or damaged hepatic
 CC tissue, means for enhancing viability and integration of hepatic
 CC tissue and organ transplants, and means for correcting liver function
 CC deficiencies, including means for enhancing diminished liver function
 CC due to tissue injury or disease.
 CC Use of BMP3(fx), BMP5(fx) and BMP6(fx) are included in the
 CC disclosure.
 SQ Sequence 402 AA;

Query Match 1.5%; Score 125; DB 10; Length 402;
 Best Local Similarity 39.7%; Pred. No. 1.45e+00;
 Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggglrppp-gc-pqrrlg-arerrdvqrellavlgpr-prprappaas 67
 QY 172 GPPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRGGASR 230
 Db 68 rlp 70
 QY 231 SLP 233

RESULT 4
 ID R54937 standard; Protein; 402 AA.
 AC R54937;
 DT 15-OCT-1994 (first entry)
 DE Osteogenic protein hOP2-PP.
 KW Morphogenic protein; hOP-2-PP; OP-2; hOP2; hOP-2;
 KW tissue morphogenesis; osteogenic protein.
 OS Homo sapiens.
 PN WO9410203-A.
 PD 11-MAY-1994.
 PF 02-NOV-1993; U10520.
 PR 03-NOV-1992; US-971091.
 PR 04-MAR-1993; US-029335.
 PR 31-MAR-1993; US-040510.
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;
 PI Pang RH, Rueger DC;
 PI WPI: 94-167392/20.
 DR N-PSDB; Q65393.
 PT A morphogenically active protein hOP-3 - for inducing tissue

```

PT morphogenesis in mammals
PS Disclosure; Page 129-131; 164pp; English.
CC A novel mouse morphogenic protein, OP3, has the sequence given in
CC R54934, and is encoded by cDNA of sequence Q65390. cDNA and protein
CC sequences were also provided for human osteogenic protein Op1
CC (Q65391, R54935), mouse Op1 (Q65392, R54936), human Op2 (Q65393,
CC R54937) and mouse Op2 (Q65394, R54938), as well as the genomic DNA
CC sequence of human Op2 (Q65395). Generic sequences given in R54939-
CC 40 accommodate homologies between Op1, Op2, Op3 and other morphogen
CC protein family members.
SQ Sequence 402 AA;

Query Match 1.58; Score 125; DB 10; Length 402;
Best Local Similarity 39.7%; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRRGGASR 230
68 rlp 70
231 SLP 233

RESULT 5
ID R60578 standard; Protein; 402 AA.
AC R60578;
DT 30-MAR-1995 (first entry)
DE Osteogenic protein OP2.
KW Osteogenic protein.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 239..242
FT /note="proteolytic cleavage site"
FT cleavage_site 260..263
FT /note="proteolytic cleavage site"
PN WO9420539-A.
PD 15-SEP-1994.
PF 04-MAR-1994; U02335.
PR 04-MAR-1993; US-027070.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Jones WK, Oppermann H, Ozkaynak E, Rueger DC, Sampath KT;
PI Tucker RF;
DR WPI: 94-302971/37.
DR P-PSDB; Q71426.
PT Binding partners, esp. antibodies, specific for different forms
PT of osteogenic protein - for differentiating between mature and
PT soluble complexed forms of the protein in culture media or serum.
PS Disclosure; Page 50-52; 70pp; English.
CC The osteogenic protein is produced recombinantly in mammalian cell
CC cultures, and may be provided to a site for bone induction in a
CC mammal with a suitable matrix to allow infiltration, proliferation
CC and differentiation of migrating progenitor cells.
SQ Sequence 402 AA;

Query Match 1.58; Score 125; DB 11; Length 402;
Best Local Similarity 39.7%; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRRGGASR 230
68 rlp 70
231 SLP 233

RESULT 6
ID R57973 standard; Protein; 402 AA.
AC R57973;
DT 11-OCT-1994 (first entry)

```

```

DE Human OP-2.
KW OP-1; OP-2; CBMP2; Vgl(fx); Vgr(fx); DPP(fx);
KW GDF-1(fx); 60A(fx); BMP3(fx); BMP5(fx); BMP6(fx);
KW tooth socket; alveolus; osteogenic protein; morphogen;
KW morphogenic protein; periodontal tissue; regeneration;
KW tooth implant; integration; inhibition.
OS Homo sapiens.
PN WO9406399-A.
PD 31-MAR-1994.
PF 15-SEP-1993; U08742.
PR 15-SEP-1992; US-945285.
PR 04-MAR-1993; US-029335.
PR 31-MAR-1993; US-040510.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;
PI Pang RHL, Rueger DC, Smart JE;
DR WPI: 94-118107/14.
DR N-PSDB; Q67313.
PT Morphogen-induced periodontal tissue regeneration - used in
PT integrating as implanted tooth in tooth socket or to inhibit
PT tissue loss associated with periodontal disease or injury
PS Claim 28-29; Page 96-98; 132pp; English.
CC Morphogens comprising an amino acid sequence sharing at least
CC 70% homology with OP-1, OP-2, CBMP2, Vgl(fx), Vgr(fx), DPP(fx),
CC GDF-1(fx), 60A(fx) and at least 80% homology with BMP3(fx),
CC BMP5(fx) and BMP6(fx) are useful for integrating an implanted
CC tooth in a tooth socket and for inhibiting tissue loss associated
CC with periodontal disease or injury.
SQ Sequence 402 AA;

Query Match 1.5%; Score 125; DB 9; Length 402;
Best Local Similarity 39.7%; Pred. No. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 glalcalggggprrpp-gc-pqrrlg-arerrdvqrellavlgpgr-prprappaas 67
QY 172 GPLYQLGAATQA-RPPPHASGPRRLGCCRANWHSVREAGVPLGLPAPGARRRGGASR 230
68 rlp 70
231 SLP 233

RESULT 7
ID R27291 standard; Protein; 402 AA.
AC R27291;
DT 26-FEB-1993 (first entry)
DE Human osteogenic protein hOP2.
KW Morphogen; morphogenic protein.
OS Homo sapiens.
PN WO9215323-A.
PD 17-SEP-1992.
PF 11-MAR-1992; U01968.
PR 11-MAR-1991; US-667274.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Cohen CM, Kuberampath T, Oppermann H, Pang RHL, Rueger DC;
DR WPI: 92-331475/40.
DR N-PSDB; Q28737.
PT Compens. for increasing progenitor cell population - contain a
PT morphogen to induce proliferation, useful for inhibiting
PT neoplastic growth, inducing tissue repair and in diagnosis of
PT tissue dysfunction
PS Disclosure; Page 93-95; 132pp; English.
CC Mature hOP2 is one of the preferred known morphogens which can be
CC used in the manufacture of pharmaceuticals for inducing non-
CC chondrogenic mammalian tissue growth, progenitor cell proliferation
CC and hepatic tissue growth and for maintaining the phenotypic
CC expression of differentiated cells in a mammal. Morphogenic
CC compositions of the invention can also be used to treat blood
CC disorders and impaired or lost immune function. Morphogens sharing
CC at least 70% homology with hOP2 are included.
SQ Sequence 402 AA;

```

21-FEB-1992; 841646.
PF PR 21-FEB-1992; US-841646.
PR PR 04-MAR-1993; US-027070.
PR PR 04-MAR-1994; US-206864.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Jones WK, Oppermann H, Ozkaynak E, Rueger DC, Sampath KT;
PI Tucker RF;
DI WPI; 97-178399/16.
DR N-PSDB; T73208.
PT Antibody specific for soluble form of osteogenic protein - for
PT quality control and diagnostic use
PT Disclosure; Column 25-28; 20pp; English.
PS This is the amino acid sequence of the human osteogenic protein 2 (OP2)
CC derived from hippocampal tissue. The invention relates to a novel
CC soluble form of this protein and of OP1 (W16365), and especially to
CC ligands binding these proteins, e.g. poly- or monoclonal antibodies.
CC The ligands are especially able to differentiate between the soluble and
CC mature forms of these proteins. The soluble forms of the proteins
CC preferably comprises the C-terminal 6 or 7 Cys residues (i.e. for OP1
CC residues 335-431 (6 Cys) or 330-341 (7 Cys)). The soluble complex
CC comprises a pro domain of the protein non-covalently linked to a dimeric
CC form of the osteogenic protein which containing at least the soluble part
CC of OP1 and another osteogenic protein e.g. OP2, bone morphogenetic
CC protein (BMP)-2, -3, -4, -5, -6 or -9. The ligand may be used to monitor
CC the purity of therapeutic osteogenic protein preparations and for
CC diagnostic purposes.
SQ Sequence 402 AA;

Query Match 1.5%; Score 125; DB 22; Length 402;
Best Local Similarity 39.7%; Pred. NO. 1.45e+00;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 gltalcaggggpglppp-gc-pqrrlg-arerrdvqreilavlgpr-prprappaas 67
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 172 GPPLYQLGQAATCA-RPPPHASGFRRLGCERAWNSVREAGVPLGLPAGARRGGCSASR 230

Db 68 rlp 70
||
QY 231 SLP 233

RESULT 10
ID R44759 standard; Protein; 402 AA.
AC R44759;
DE 06-JUN-1994 (first entry)
DT Human osteogenic protein OP2.
KW Osteogenic protein; bone; cartilage; matrix; osteoarthritis;
KW repair; vascularisation; mineralisation; differentiation.
OS Homo sapiens.
PN US5266683-A.
PD 30-NOV-1993.
PF 08-APR-1988; 179406.
PF 08-APR-1988; US-179406.
PR 15-AUG-1988; US-232630.
PR 23-FEB-1989; US-315342.
PR 17-OCT-1989; US-422613.
PR 17-OCT-1989; US-422699.
PR 22-FEB-1990; US-483913.
PR 20-AUG-1990; US-569920.
PR 17-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 18-OCT-1990; US-600024.
PA 04-DEC-1990; US-621849.
PA 04-DEC-1990; US-621988.
PA 22-FEB-1991; US-660162.
PA 20-DEC-1991; US-810560.
PA 28-JAN-1992; US-827052.
PA 21-FEB-1992; US-841646.
PI Kuberasampath T, Oppermann H, Ozkaynak E, Pang RHL;
PI Rueger DC;
DR WPI; 93-395405/49.
DR N-PSDB; Q53155.
PT New pure mammalian osteogenic proteins - induce cartilage and

	WU0238;	
AC	DT 21-NOV-1996 (first entry)	
DE	Human osteogenic protein OP-2.	
KW	Morphogen; osteogenic protein; dentine; tooth decay; caries;	
KW	Morphogenesis; odontoblast; Op-2.	
OS	Homo sapiens.	
FH	Key Location/Qualifiers	
FT	peptide 1..17	/label= Sig_peptide
FT	region 18..263	/label= Pro_region
FT	protein 264..402	/label= Mat_protein
FT	domain 301..402	/label= 7-Cys_C-terminal_domain
PN	WO9626737-A1.	
PD	06-SEP-1996.	
PB	14-FEB-1996; UO2169.	
PR	01-MAR-1995; US-396930.	
PA	(CREA-) CREATIVE BIOMOLECULES INC.	
PI	Charette MF, Rutherford RB,	
DR	WPL: 96-412583/41.	
DR	N-PSTDB; T33443.	
PT	Use of morphogen(s), e.g. osteogenic proteins, on dental surfaces	
PT	- for inducing dentine morphogenesis, desensitising teeth or sealing	
PT	tooth cavities	
PS	Disclosure: Page 58-59; 106pp; English.	
CC	Human hippocampus full-length osteogenic protein OP-2 (W00238)	
CC	includes a pro-sequence and the morphologically active mature	
CC	protein sequence (see also W00223) that includes a 7-Cys C-terminal	
CC	domain. OP-2 can be expressed from intact or truncated cDNA	
CC	(T33443) in prokaryotic or eukaryotic host cells. Mature OP-2 and	
CC	other morphogens (see also W00221-35), partic. human OP-1 (W00221),	
CC	can be used to induce dentine morphogenesis, to seal dental cavities	
CC	and to desensitize teeth to pressure and/or temp.	
SQ	Sequence 402 AA;	
	Query Match 1.5%; Score 125; DB 19; Length 402;	
	Best Local Similarity 39.7%; Pred. No. 1.45e+00;	
	Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps	
Dd	12 glalcalgggpglrrpp-gc-pqrlig-arerrdvarellavlgpr-prprappaas 67 : :: : : :: : :	:: : : :
Qy	172 GPLLYQLGAATQA-RPPPHASGRRRLGCERANNHSYREAGVLGLPAPGARRRGGSAZR 233 : :: : : :: : : :: : : :	:: : : :
Dd	68 rlp 70	:
Qy	: : 231 SLP 233	:
RESULT	13	
ID	R47292 standard; Protein; 402 AA.	
AC	R47292;	
DE	02-SEP-1994 (first entry)	
KE	hop2.	
KD	Human; hippocampus; osteogenic protein; OP-1; hop-1; mature; injury;	
KW	survival; neural cell; morphogen; chemical; mechanical; neuropathy;	
KW	transected nerves; demyelinated cell; toxin; ethanol; parkinsons;	
KW	Alzheimers; Huntingtons chorea; amyotrophic lateral sclerosis;	
KW	multiple sclerosis; neoplastic lesion; central nervous system; CNS;	
KW	retinoblastoma; glial cell neoplasm; redifferentiation; neuroblastom;	
KW	peripheral nervous system; neurite; outgrowth; cell aggregation;	
KW	cell adhesion; axonal; nerve regeneration; vascularisation;	
KW	myelin sheath.	
OS	Homo sapiens.	
PN	WO9403200-A.	
PD	17-FEB-1994.	
PR	29-JUL-1993; UO7189.	
PR	31-JUL-1992; US-922813.	
PR	04-MAR-1993; US-029335.	
PR	31-MAR-1993; US-040510.	
PA	(CREA-) CREATIVE BIOMOLECULES INC.	
PI	Jones WK, Kuberasampah T, Oppermann H, Ozkaynak E;	

Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;
 Db 12 glalcalggggpdlrppp-gc-pqrrlg-arerrdvqrellavlglpgr-prprappaas 67
 Qy 172 GPPLYQLGAATQA-RPPPHASGPRRLGCRANWHSYREAGVPLGLFPAGARRRGGSASR 230
 Db 68 rlp 70
 Qy 231 SLP 233

Search completed: Thu Jul 30 14:44:41 1998
 Job time : 102 secs.

WIRE (TW)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 30 14:44:58 1998; Maspar time 42.67 Seconds
969.188 Million cell updates/sec

Linear output not generated.

Title: >US-08-912-951-2
Description: (1-1132) from US08912951.pep
Perfect Score: 8465
Sequence: 1 MPRAPRCRAVRLRLSHYRE.....TALRAANPALPDKFTILD 1132

Scoring table: PAM 150
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir56
1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 53.528; Variance 118.315; scale 0.452

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result	Query	Match	Length	ID	Description	Pred. No.	
1	314	3.7	884	2	S3396	hypothetical protein	5.31e-30
2	162	1.9	660	1	QBE3	BHLF1 protein - human	1.01e-06
3	153	1.8	504	2	JC1306	virion protein homolo	1.61e-05
4	141	1.7	276	2	B38965	hypothetical protein	5.73e-04
5	142	1.7	310	1	PIHUSD	salivary proline-rich	4.28e-04
6	136	1.6	240	2	B24264	proline-rich protein	2.43e-03
7	136	1.6	316	2	S16681	homeotic protein - hu	2.43e-03
8	132	1.6	317	2	A28996	proline-rich protein	7.54e-03
9	137	1.6	628	2	S01955	hypothetical protein	1.82e-03
10	132	1.6	924	2	S27923	gene LF3 protein - hu	7.54e-03
11	125	1.5	227	2	C29149	proline-rich protein	5.22e-02
12	123	1.5	240	2	A24264	proline-rich protein	8.97e-02
13	123	1.5	256	2	A00533	tumor-associated anti	8.97e-02
14	129	1.5	264	2	D34768	ORF4 protein - Orf vi	1.74e-02
15	131	1.5	300	2	S19560	proline-rich protein	9.98e-03
16	125	1.5	302	2	S11790	modulation protein no	5.22e-02
17	129	1.5	309	2	S10889	proline-rich protein	1.74e-02
18	125	1.5	381	2	S16506	hypothetical protein	5.22e-02
19	125	1.5	402	2	A45056	prepro osteogenic pro	5.22e-02
20	123	1.5	403	2	S52796	prpL2 protein - human	8.97e-02
21	123	1.5	439	2	S51939	chitinase (EC 3.2.1.1	8.97e-02
22	124	1.5	464	2	S22697	extensin - Volvox car	6.85e-02
23	127	1.5	566	2	S22933	testis-specific prote	3.02e-02

24 131 1.5 1106 2 JQ0405 hypothetical 119.5K p 9.98e-03
25 117 1.4 204 2 A39066 proline-rich protein 4.39e-01
26 115 1.4 260 2 S23773 proline-rich protein 7.36e-01
27 116 1.4 270 2 S43361 miag protein - Salmon 5.68e-01
28 116 1.4 295 2 B48013 proline-rich proteogl 5.68e-01
29 118 1.4 322 1 S00054 modulation protein no 3.38e-01
30 115 1.4 338 2 JC5707 HYA22 protein - human 7.36e-01
31 119 1.4 347 2 S10571 epithelial tumor anti 2.60e-01
32 121 1.4 358 1 WMBE38 infected cell protein 1.53e-01
33 115 1.4 392 1 PIHUB6 salivary proline-rich 7.36e-01
34 118 1.4 431 2 S09824 hypothetical protein 3.38e-01
35 119 1.4 515 2 S10572 epithelial tumor anti 2.60e-01
36 115 1.4 3149 1 QBE8 BPLF1 protein - human 7.36e-01
37 114 1.3 202 2 B36795 hypothetical protein 9.50e-01
38 114 1.3 301 2 E29149 proline-rich protein 9.50e-01
39 114 1.3 321 1 A28663 modulation protein no 9.50e-01
40 114 1.3 373 2 S54545 hypothetical protein 9.50e-01
41 114 1.3 408 2 A56186 cyclin E - African cl 9.50e-01
42 114 1.3 705 2 A35363 synapsin I splice for 9.50e-01
43 113 1.3 1255 2 B35175 episialin B - human 1.23e+00
44 113 1.3 1264 2 A35175 episialin A - human 1.23e+00
45 113 1.3 1295 2 A35886 polymorphic epithelia 1.23e+00

ALIGNMENTS

RESULT 1
ENTRY S53396 #type complete
TITLE hypothetical protein YLR318w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein L8543.12
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 06-Feb-1998
ACCESSIONS S53396
REFERENCE S53390
#authors Du, Z.
#submission submitted to the EMBL Data Library, February 1995
#description The sequence of S. cerevisiae cosmid 8543.
#accession S53396
#molecule_type DNA
#residues 1-884 #label DUZ
#cross-references EMBL:U20618; NID:g2258165; PID:g662136; MIPS:YLR318w
#experimental_source strain S288C (AB972)
GENETICS
#gene SGD:EST2
#map_position 12R
#cross-references SGD:S0004310; MIPS:YLR318w
SUMMARY #length 884 #molecular_weight 102662 #checksum 7604

Query Match 3.7%; Score 314; DB 2; Length 884;
Best Local Similarity 24.3%; Pred. No. 5.31e-30;
Matches 118; Conservative 122; Mismatches 202; Indels 44; Gaps 37;
Db 271 LSHLSRQSPKER-VLKFFIVILQKLLPQEMGSKKNGKIKNLLNLLSLPLNGVLPDS 329
Qy 448 LVQLLRQSHSPQVYGFYRACLRLDPGLMGSRNRRFLRNTKTKFSLGKHAQLSQE 507
Db 330 LLKLRKLDKFWLFTS-DIWFTKHNFNLNQALIC-FISWLFRLQIPKIQFFCYCTEIS 387
Qy 508 LTKWSVRDCAWLRSPGVCVPAEHLREILAKFLHLMMSVYVVELLRFFVTTET 567
Db 388 STVTIVYF-RHDTNKLITPFIVEYFKTY-LVE-NNV-CRNHNSYTLNFNHSMRIIPK 443
Qy 568 FQKNRLEFYRKSWSKLQSIGIRQHLKRVQLRELSEAEVQHQRE-ARPAALLTSRLRFIPK 626
Db 444 KSNNEFRILAPCGADDEEFTIYKENHK-NA-IQPTOKILEYLNKRPTSF-TKIYSPT 500
Qy 627 -PDGLRPIVNDY-VVGARTFRERKRAERLTSRVKALFSLVNYERARRPGLLGASVLGD 684
Db 501 QIADRI-KEFKORLKKFNVLPELYEMKFDVKSCYDSITPRMECHMR-ILKDALKNENGFF 558
Qy 685 DI-HRAWTFVLVR-RA-QDPPPELIFYKVDVTGAYDTIPQ-DRUTEVIASIKRPTONTYC 740


```

REFERENCE      JCL1306
#authors       Carpenter, D.E.; Misra, V.
#journal       Gene (1992) 119:259-263
#title         Sequences of the bovine herpesvirus 1 homologue of herpes simplex virus type-1 alpha-trans-inducing factor (UL48).
#accession     JCL1306
##molecule_type DNA
##residues     1-504 #label CAR
##cross-references EMBL:Z11610; NID:g1065725; PID:e264419; PID:g1065726
COMMENT        This protein interacts with cellular transcription factors to transactivate immediate early viral genes.

GENETICS
#map_position  0.07-0.086
KEYWORDS       DNA binding; transcription regulation
SUMMARY        length 504 #molecular-weight 54028 #checksum 8743

Query Match          1.8%; Score 153; DB 2; Length 504;
Best Local Similarity 31.3%; Pred.No. 1.61e-05;
Matches              46; Conservative 33; Mismatches 56; Indels 12; Gaps 11

Db    351 APAEAGGWRRSSTTRGAARATGTIRLQRPCCGPRRAKC-CRATP-RORLR--ARGE 406
      :|: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |::
QY    186 PPHASGPRRLGCERAWNSVREAGVGLGPAPCARRRGGSSASRLPLPKRPREGAAPE 245

Db    407 PRUTS-GSGAFSQ-GRRGVRCVLGWACKARGPGARGGPGSPVRSGIGLSR-ARGSPGP 463
      |: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |::
QY    246 PERTPVQGSAHPHRTGPSDR-GF-CVV-S-PARPAEATSLEGALSGTRHSHPSVGR 301

Db    464 GPACGGPSRARGRRRASPNP-FGGT 489
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY    302 QHAGPPSTSRPPRPMDTPCPPVIAET 328

RESULT      4
ENTRY       B38965 #type complete
TITLE       hypothetical protein B (Insertion sequence IS1222) -
            Enterobacter agglomerans
ORGANISM    03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change
DATE        09-Sep-1997
ACCSSIONS   B38965; S42922
REFERENCE    B38965
#authors     Steibl, H.D.; Lewecke, F.M.
#journal     Gene (1995) 156:37-42
#title       IS1222: analysis and distribution of a new insertion sequence
            in Enterobacter agglomerans 339.
#accession   B38965
#status      preliminary
##molecule_type DNA
##residues   1-276 #label STE
##cross-references GB:X78052; NID:g459246; PID:g459248
SUMMARY      length 276 #molecular-weight 31718 #checksum 3927

Query Match          1.7%; Score 141; DB 2; Length 276;
Best Local Similarity 33.7%; Pred.No. 5.73e-04;
Matches              30; Conservative 23; Mismatches 30; Indels 6; Gaps 6;

Db    42 ITELALERR-FCYRR-INOLLRRSLGHNVHKRVLYHLSCGLGVKRRR-RKGILATERL 98
      :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY    563 VTETTQKNRLFYKSWSKQSITGIROHLKRV-QLRLESEAVRQHREARPALLTSRL 621

Db    99 PLL-RPAAPNLTWSDDFVMDALATGRRIK 126
      ::: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |:: |: |::
QY    622 RFIPKPDGLRPVNMDYVVGGA-RTFREX 649

RESULT      5
ENTRY       PIHUSD #type complete
TITLE       salivary proline-rich glycoprotein precursor PRB4 (large allele) - human
CONTAINS    basic proline-rich protein IB-5; proline-rich peptide P-D
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        19-Feb-1984 #sequence_revision 12-Apr-1996 #text_change

```

```

ACCESSIONS      05-Sep-1997
S03176; S03175; S10890; D25372; E38355; A03295; A61294;
REFERENCE
#authors      Lyons, K.M.; Stein, J.H.; Smithies, O.
#journal      Genetics (1988) 120:267-278
#title       Length polymorphisms in human proline-rich protein genes
              generated by intragenic unequal crossing over.
#cross-references MUID:89121440
#accession    S03176
#molecule_type DNA
#status      translation not shown
#residues    35-310 ##label LY1
#cross-references EMBL:X07715
#note       large allele
#accession    S03175
#molecule_type DNA
#status      translation not shown
#residues    35-36, 'E', 38-112, 155-310 ##label LY2
#cross-references EMBL:X07704
#note       medium allele
#accession    S10890
#molecule_type DNA
#status      preliminary; translation not shown
#residues    1-38, 60-112, 'T', 114-115, 'P', 117-121, 185-271, 'A', 273-310
              ##label LY3
#cross-references EMBL:X07882; NID:g35647; PID:g296670
REFERENCE
#authors      Mada, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
#journal      J. Biol. Chem. (1985) 260:11123-11130
#title       Differential RNA splicing and post-translational cleavages in
              the human salivary proline-rich protein gene system.
#cross-references MUID:85289325
#accession    D25372
#molecule_type mRNA
#residues    1-36, 'E', 38-112, 'T', 114-115, 'P', 117-121, 185-271, 'A',
              273-310 ##label MAE
REFERENCE
#authors      Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
#journal      Biochemistry (1991) 30:3351-3356
#title       Basic proline-rich proteins from human parotid saliva:
              relationships of the covalent structures of ten proteins
              from a single individual.
#cross-references MUID:91190884
#accession    E38355
#molecule_type protein
#residues    241-254, 'KN', 257-310 ##label KAU
REFERENCE
#authors      Saitoh, E.; Isemura, S.; Sanada, K.
#journal      J. Biochem. (1983) 93:495-502
#title       Complete amino acid sequence of a basic proline-rich peptide,
              P-D, from human parotid saliva.
#cross-references MUID:83186122
#accession    A03295
#molecule_type protein
#residues    241-310 ##label SAI
REFERENCE
#authors      Shimomura, H.; Kanai, Y.; Sanada, K.
#journal      J. Biochem. (1983) 93:857-863
#title       Amino acid sequences of glycopeptides obtained from basic
              proline-rich glycoprotein of human parotid saliva.
#accession    A61294
#molecule_type protein
#residues    54-57, 'E', 59-73, 'R', 82-101 ##label SHI
REFERENCE
#authors      Charlton, A.J.; Baxter, N.J.; Lilley, T.H.; Haslam, E.;
              McDonald, C.J.; Williamson, M.P.
#journal      FEBS Lett. (1996) 382:289-292
#title       Tannin interactions with a full-length human salivary
              proline-rich protein display a stronger affinity than with
              single proline-rich repeats.
#accession    S62891
#molecule_type protein

```

```

#residues      241-252 ##label CHA
#note         amino end of peptide designated basic proline-rich
              protein IB-5
#note         it is unclear from the peptide sequence whether this is
              a product of the PRB2 (PIR:PIHUPF) or PRB4 (this
              entry) gene
GENETICS
#gene         GDB:PRB4
#cross-references GDB:119514; OMIM:180990
#map_position 12p13.2-12p13.2
#introns      22/1: 34/1
#note         the list of introns may be incomplete
CLASSIFICATION #superfamily proline-rich protein
KEYWORDS       glycoprotein; saliva; tandem repeat
FEATURE
1-16          #domain signal sequence #status predicted #label SIG\
241-310       #product proline-rich peptide P-D #status experimental
66,87,171     #label MAR\
108,150,192,213,
234           #binding_site carbohydrate (Asn) (covalent) #status
              experimental\
SUMMARY       #length 310 #molecular-weight 31351 #checksum 3960
              Predicted
              Query Match 1.7%; Score 142; DB 1; Length 310;
              Best Local Similarity 28.2%; Pred. No. 4.28e-04;
              Matches 40; Conservative 41; Mismatches 52; Indels 9; Gaps 9;
Db 168 QGNQSQG-PPPHPGKERPPP-QGG-NQSHRPPPP-GKPER-PPPGGNSQG-P-PP 220
QY 177 QLGAAATQARPPPHASGPRRLGGERANVHVSREAGVPLGLPAPGARRRGGASRSLLPK 236
Db 221 HPGKPEGPPOEGNKSRSAPSGKPGQPGQEGNKQPGPPPKPGQPPPPGPNQPOQ 280
QY 237 RPRGAAPERPTVGGQSWAHFGRTRGSDR-GFCVVSAPPAE-EATSLGALSGRH 294
Db 281 APPAGKPGPPPPPPGGRPPRP 302
QY 295 SHFSVGRQHAGPPSTRPPRP 316
RESULT 6
ENTRY       B24264 #type fragment
TITLE       proline-rich protein MP3 - mouse (fragment)
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change
              03-May-1996
ACCESSIONS  B24264
REFERENCE    A92508
#authors     Ann, D.K.; Carlson, D.M.
#journal     J. Biol. Chem. (1985) 260:15863-15872
#title       The structure and organization of a proline-rich protein gene
              of a mouse multigene family.
#cross-references MUID:86059475
#accession   B24264
#molecule_type DNA
#residues    1-240 ##label ANN
CLASSIFICATION #superfamily proline-rich protein
SUMMARY       #length 240 #checksum 5152
              Query Match 1.6%; Score 136; DB 2; Length 240;
              Best Local Similarity 27.2%; Pred. No. 2.43e-03;
              Matches 41; Conservative 38; Mismatches 65; Indels 7; Gaps 6;
Db 24 VNGSQGPPPPGPGQPPRPQPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 83
QY 178 LGAAATQARPPPHASGPRRLGGERANVHVSREAGVPLGLPAPGARRRGGASRSLLPKR 237
Db 84 PQGPPPPGPGQPPRP-PQGPPPPGPGPPPPGPPPPGPPGPPPPGPPGPPHRLRPTQG 142
QY 238 PRGANPE-PERTVGGQSWAHFG-RTRGSDRGFCVVSAPPAEATSLGALSGRH 294

```

```

Db 143 PPPGGPQPRPPQGGPPTGTQPRPTQGP-PP 172
QY 295 SHPSVGRQHA--GPPSTRPPRPWDTPCPP 323

RESULT 7
ENTRY title homeotic protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 17-Oct-1997
ACCESSIONS S16681
REFERENCE S16681
#authors Deguchi, Y.; Kehrl, J.H.
#journal Nucleic Acids Res. (1991) 19:3742
#title Nucleotide sequence of a novel diverged human homeobox gene
#cross-references MUID:91305125
#accession S16681
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type mRNA
#residues 1-316 #label DEG
#cross-references EMBL:X56537
#note the nucleotide sequence was submitted to the EMBL Data Library, January 1991
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation
FEATURE 73-127
#domain homeobox homology #label HOX
SUMMARY #length 316 #molecular-weight 34713 #checksum 3095

Query Match 1.6%; Score 136; DB 2; Length 316;
Best Local Similarity 30.9%; Pred. No. 2.43e-03;
Matches 21; Conservative 20; Mismatches 24; Indels 3; Gaps 3;

Db 202 RPNSTASSCSVGTGTAPRGPAT-SPRRSRPAAQDRSPARAPGAASPGGAGWTH 260
QY 201 RANWHSVREAGV-PLGLPAPGARRGGASRSRLPLKPRRG-AAPEPRTVPVGGSWAH 258
Db 261 PARPREQA 268
QY 259 PGRTRGPS 266

RESULT 8
ENTRY title proline-rich protein M14 precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 20-Mar-1998
ACCESSIONS A28996
REFERENCE A28996
#authors Ann, D.K.; Smith, K.; Carlson, D.M.
#journal J. Biol. Chem. (1988) 263:10887-10893
#title Molecular evolution of the mouse proline-rich protein multigene family. Insertion of a long interspersed repeated DNA element.
#cross-references MUID:88273214
#accession A28996
#molecule_type DNA
#residues 1-317 #label ANN
#cross-references GB:M23236; GB:J03891; NID:g200535; PID:g567232

GENETICS 22/1
#introns
CLASSIFICATION #superfamily proline-rich protein
KEYWORDS saliva
FEATURE 1-15
#domain signal sequence #status predicted #label SIG\
#product proline-rich protein M14 #status predicted #label MAR
#length 317 #molecular-weight 31719 #checksum 8454
SUMMARY

Query Match 1.6%; Score 137; DB 2; Length 628;
Best Local Similarity 25.0%; Pred. No. 1.82e-03;
Matches 73; Conservative 64; Mismatches 134; Indels 21; Gaps 20;

Db 157 GPVLTETKPTSVRQPSATRGPSFRPILLPKV-VH-VHDDPPHSSLRPRGSRSLQPT 214
QY 172 GPPLYQLGAATQARPPHAS-GPR-RRIGCERAWNHVSREAGVPLGLPAPGAR-RR- 227
Db 215 VRRPLAPNOFHSPPQPLSDPDGILGPRPLAPHSTRDPPRPPI-TGPGSN-THDLRPL 272
QY 228 ASRSLPLKPRRGAAPEPERTPVG-QGSA-HGRTRGPSDRGFCVVSAPARAEATSL 285
Db 273 SVLPRTSPRRGLLPNRRHTSTGHIPTTTSRPTGPPSLRQRPVH-LYOSSPTTPNRP 331
QY 286 EG-ALSGTRSHSPVGRQHAGPSTSRPPRPMDTPCPVYAEKHELY-SSGDEQLRP 343
Db 332 SSIRKDALLOTGRLGHLERLGGQANLRTSERSPTTKRLRPSSEPNRLPKPLPEATLP 391
QY 344 SFLL-SSL-R--PSLTGARLRLVETIFLGRPMWPGTGRPLPQRYQWQMRPLFLLIG- 398
Db 392 SYRHRPYPPLPNPPAALPSIATVTSRRGIHSHSLKGLPK-EGAPPPPRRL 442
QY 399 NHAQC-PYGVLLKTHCPLRA-AVTPAAGVCAREKPGSQSVAAPAEEDTPRRL 448

RESULT 10
ENTRY title gene LF3 protein - human herpesvirus 4
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 08-Sep-1997
ACCESSIONS S27923
REFERENCE S27923

```

```

Query Match 1.6%; Score 132; DB 2; Length 317;
Best Local Similarity 26.4%; Pred. No. 7.54e-03;
Matches 39; Conservative 38; Mismatches 64; Indels 7; Gaps 6;

Db 37 SGSPRPVNGSQGPPPPGQPRP-PQGPVPPGQPRPQGPVPPGQPRP-PQGP 94
QY 180 AATQARPPPHAS--GPRRLGCE-RANWHSVREAGVPLGLPAPGARRGSGASLSLPK 236
Db 95 PPGGQPRPPGPPPPGG-PQPRPQGPVPPGQPRPQGPVPPGQPRPQGPVPP 153
QY 237 RPRGAAPERTPVGQSWAHGRTGRGSDRGFCVVSAPAR-PAEATSLGSLGSTRHS 295
Db 154 GGPQPRPPQGPVPPGQPRPQGPVPP 181
QY 296 HPSVGRQHAGPSTSRPPRPWDTPCPP 323

RESULT 9
ENTRY title #type complete
ORGANISM #formal_name turnip yellow mosaic virus, TYMV
DATE 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 26-May-1995
ACCESSIONS S01955
REFERENCE S01955
#authors Morch, M.D.; Boyer, J.C.; Haenni, A.L.
#journal Nucleic Acids Res. (1988) 16:6157-6173
#title Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA.
#cross-references MUID:88289359
#accession S01955
#status preliminary
#molecule_type genomic RNA
#residues 1-628 #label MOR
#cross-references EMBL:X07441
#note the authors translated the codon ACG for residue 459 as U
SUMMARY #length 628 #molecular-weight 69194 #checksum 7569

Query Match 1.6%; Score 137; DB 2; Length 628;
Best Local Similarity 25.0%; Pred. No. 1.82e-03;
Matches 73; Conservative 64; Mismatches 134; Indels 21; Gaps 20;

Db 157 GPVLTETKPTSVRQPSATRGPSFRPILLPKV-VH-VHDDPPHSSLRPRGSRSLQPT 214
QY 172 GPPLYQLGAATQARPPHAS-GPR-RRIGCERAWNHVSREAGVPLGLPAPGAR-RR- 227
Db 215 VRRPLAPNOFHSPPQPLSDPDGILGPRPLAPHSTRDPPRPPI-TGPGSN-THDLRPL 272
QY 228 ASRSLPLKPRRGAAPEPERTPVG-QGSA-HGRTRGPSDRGFCVVSAPARAEATSL 285
Db 273 SVLPRTSPRRGLLPNRRHTSTGHIPTTTSRPTGPPSLRQRPVH-LYOSSPTTPNRP 331
QY 286 EG-ALSGTRSHSPVGRQHAGPSTSRPPRPMDTPCPVYAEKHELY-SSGDEQLRP 343
Db 332 SSIRKDALLOTGRLGHLERLGGQANLRTSERSPTTKRLRPSSEPNRLPKPLPEATLP 391
QY 344 SFLL-SSL-R--PSLTGARLRLVETIFLGRPMWPGTGRPLPQRYQWQMRPLFLLIG- 398
Db 392 SYRHRPYPPLPNPPAALPSIATVTSRRGIHSHSLKGLPK-EGAPPPPRRL 442
QY 399 NHAQC-PYGVLLKTHCPLRA-AVTPAAGVCAREKPGSQSVAAPAEEDTPRRL 448

RESULT 10
ENTRY title #type complete
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 08-Sep-1997
ACCESSIONS S27923
REFERENCE S27923

```

```

#authors Parker, B.D.; Bankier, A.; Satchwell, S.; Barrell, B.;
Farrell, P.J.
#submission submitted to the EMBL Data Library, August 1990
#description Sequence and transcription of Raji Epstein-Barr virus DNA
spanning the B95-8 deletion region.
#accession S27923
##status preliminary
##molecule_type DNA
##residues 1-924 ##label PAR
##cross-references EMBL:M35547; NID:g330420; PID:g330421
SUMMARY #length 924 #molecular-weight 94304 #checksum 8997

Query Match 1.68; Score 132; DB 2; Length 924;
Best Local Similarity 28.28; Pred.No.7.54e-03;
Matches 37; Conservative 33; Mismatches 52; Indels 9; Gaps 9;

Db 38 AAPRAGPEPTRLQATPRRSGAADPADPVGHPPAA-PRAPGPEPTRLQ-PATPRRSGA 95
QY 184 ARPPPHASGPRRLGGERAWNHVSRGAGVPLGLPAGARRRGSASRSLPLRRPRRGA 243
96 ADPA-DPVGHPPA-A-P-RAPGPEPTRLQATPRRSGAADPADPVGHPPAAPRAPGPEPTR 151
QY 244 PEPERTPVGGSWAHGTRGPDSDRG-FCVVSARP-AEEATSLGALSGRHSHP-SVG 300
Db 152 RLQATPRRSG 162
QY 301 ROHHAGPPSTS 311

RESULT 11
ENTRY C29149 #type fragment
TITLE proline-rich protein - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change
20-Mar-1998
ACCESSION C29149
REFERENCE C29149
AUTHOR Clements, S.; Mehansho, H.; Carlson, D.M.
J. Biol. Chem. (1985) 260:13471-13477
#journal Novel multigene families encoding highly repetitive peptide
sequences. Sequence analyses of rat and mouse proline-rich
protein cDNAs.
#cross-references MUID:86033799
#contents Clone PUMP40
#accession C29149
##molecule_type mRNA
##residues 1-227 ##label CLE
##cross-references GB:M11902; NID:g200544; PID:g200545
#classification #superfamily proline-rich protein
#length 227 #checksum 1378
Query Match 1.58; Score 125; DB 2; Length 227;
Best Local Similarity 27.38; Pred.No.5.22e-02;
Matches 41; Conservative 34; Mismatches 67; Indels 8; Gaps 8;

Db 49 QOGPPPPGGQPRPPQGGPPGQPRP-PQGPPPGQPRPPQGGPPPGGQPRP-PQ 106
QY 177 QLGAAQTARPPHAS-GPRRLGCE-RAWNHVSREAGVPLGLPAGARRRGSASRSLPL 234
Db 107 GPPPPGGQPRPPQGGPPPGG-QPDRPPQGGPPPGGQPRPPQGGPPPGGQPRPP 165
QY 235 PKRPRRGAAPERTPVGGSWAHPGTRGPSDRGFCVVVSARP-PAEATSLGALSCTR 293
Db 166 PRAPGQPRPPQ-GPPPTGQPRPTQGP-PP 193
QY 294 HSHPSVGRQHAGPPSTSRPRRPWDTPC 323

RESULT 12
ENTRY A24264 #type fragment
TITLE proline-rich protein MP2 - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change

```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
MPArch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 30 14:47:32 1998; MasPar time 28.16 Seconds
1008.379 Million cell updates/sec
MasPar output not generated.
```

```
>US-08-912-951-2
Title:
Description:
Perfect Score: 8465
Sequence: 1 MPRAPRGRVRLSHRYRE .....TALEAANPALPDSFKTILD 1132
```

Scoring table: PAM 150
Gap 11

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot35
1:swiss1

Statistics: Mean 55.788; Variance 96.504; scale 0.578

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Rank	Score	Query Match	Length	DB ID	Description	Pred. No.
1	162	1.9	660	1	YHL1_EBV	4.58e-09
2	150	1.8	234	1	PRP1_HUMAN	4.42e-07
3	153	1.8	504	1	ATIN_HSVBP	1.43e-07
4	142	1.7	247	1	PRPL_HUMAN	8.39e-06
5	133	1.6	247	1	PRP4_HUMAN	2.05e-04
6	137	1.6	628	1	V70K_TYRV	5.03e-05
7	125	1.5	261	1	PRP2_MOUSE	3.14e-03
8	131	1.5	296	1	PRP3_MOUSE	4.10e-04
9	125	1.5	302	1	NO3_RHLIP	3.14e-03
10	125	1.5	402	1	BMP8_HUMAN	3.14e-03
11	127	1.5	566	1	TS13_MOUSE	1.60e-03
12	116	1.4	270	1	MIAE_SALTY	5.84e-02
13	115	1.4	279	1	Y091_NPVOP	7.99e-02
14	121	1.4	322	1	NO11_RHISN	1.17e-02
15	115	1.4	331	1	PRP1_HUMAN	7.99e-02
16	118	1.4	399	1	BM8A_MOUSE	3.09e-02
17	118	1.4	431	1	U161_HCMVA	3.09e-02
18	118	1.4	514	1	VF21_PVPOV	3.09e-02
19	115	1.4	3149	1	TEGU_EBV	7.99e-02
20	107	1.3	174	1	BAR1_CHITE	9.11e-01
21	114	1.3	202	1	VG01_HSVBP	1.09e-01
22	106	1.3	233	1	YFJR_ECOLI	1.22e+00
23	106	1.3	285	1	YAFY_ECOLI	1.22e+00

ALIGNMENTS

```

RESULT 1
ID YHLL_EBV STANDARD; PRT; 660 AA.
AC P03181;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL BHLF1 PROTEIN.
OS EPSTEIN-BARR VIRUS (STRAIN 95-8) (HUMAN HERPESVIRUS 4).
SC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
NC [1]
RP PROTEIN CODING REGION.
RX MEDLINE; 84270867.
RA BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,
RA TUFFENELL P.S., BARRELL B.G.;
RL NATURE 310:207-211(1984).
DR EMBL; V01555; -; NOT_ANNOTATED_CDS.
DR PIR; A03742; Q0B83.
KW HYPOTHETICAL PROTEIN; EARLY PROTEIN; REPEAT.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 56244 MW; 372F08C5 CRC32;

Query Match 1.98; Score 162; DB 1; Length 660;
Best Local Similarity 30.8%; Pred. No. 4.58e-09;
Matches 48; Conservative 34; Mismatches 60; Indels 14; Gaps 12;

Db 251 GPPTRSGAAQ-RTHRRPGCPRSARNPCGPTWRR--R-SGAQRGHPPPGAGQRPSPG 306
QY 172 GPPLYQUGAQATQRPHPHAG-PRR-R-LGCEAWNHSVREAGVJGLPAPGA-RRRGS 227
Db 307 TGGPRAAPGAPGTAAAPGPGGGAAVSGAPHPHERGSGPADPP--AAARLPPRCQRPRLP 364
QY 228 ASRSLPLPKPRCAPEPER-TPVQGSWAHPGRTGPGSDRGFCVVSAPARPAEATSLE 286
Db 365 QDLAAQRC-PAGPPPTRSGA-AAQKTRH-RPPGCP 397
QY 287 GALSGRHSPSVGRQHHAGPPSPTRPPRWDTCP 322

RESULT 2
ID PRPM_HUMAN STANDARD; PRT; 234 AA.
AC P10161; P02813;
DT 01-MAR-1989 (REL. 10, CREATED)

```

Db 351 APAEAGGWRSSSTTRGRAARSTTGRQLRPPCCGGPRRAK-CRATP-QRLR--ARGE APO 400
 :
 QY 186 PPHASGPRRLGLCERAWNHVSREAGVPLGLPAPGARRRGGGSARSRLPLPKPRRGAAPE 245
 :
 Db 407 PRHTS-GSGAFSQ-GRPGRVCLGWACKARSGPARGGPCPSVRSGGLGLSR-ARGSPGP 463
 :
 QY 246 PERTPVQGGWAHPGTRGPSDR-GF-CVV-S-PARPAEATSLEGALSGTRHSFVSVR 301
 :
 Db 464 GPACGGPSRARGGRRRASPNP-FGGT 489
 :
 QY 302 QHHAGPSTSRPPRPWDTPCPVVAET 328
 :

```

RESULT      4
ID   FRP4_HUMAN STANDARD;          PRT;    276 AA.
AC   P10162; P02813;
DC   01-MAR-1989 (REL. 10, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DI   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   SALIVARY PROLINE-RICH PROTEIN PO (ALLELE K) (CONTAINS: PEPTIDE P-D)
GN   (FRAGMENT).
GE   GN PRB4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
NC   EUTHERIA; PRIMATES.
RN   [1]
RS   SEQUENCE FROM N.A.
RA   LYONS K.M., STEIN J.H., SMITHIES O.;
RL   GENETICS 120:255-265(1988).
RM   [2]
RR   SEQUENCE OF 207-276.
RX   MEDLINE; 83186122.
RA   SAITOH E., ISEMURA S., SANADA K.;
RL   J. BIOCHEM. 93:495-502(1983).
DR   EMBL; X07715; E4806; ALT_SEQ.
DR   PIR; A03295; PIHUSD.
DR   PIR; S03176; S03176.
DR   HSSP; P19999; 1CLG.
DR   MIM; 168730; -.
DR   MIM; 180990; -.
KW   REPEAT; PAROTID GLAND; SALIVA; MULTIGENE FAMILY.
FT   NON_TER        1
FT   CHAIN           207..276             PEPTIDE P-D.
FT   SEQUENCE        276 AA; 27816 MW; 4838945A CRC32;

Query Match       1.7%; Score 142; DB 1; Length 276;
Best Local Similarity 28.2%; Pred. No. 8.39e-06;
Matches 40; Conservative 41; Mismatches 52; Indels 9; Gaps

Db 134 QCGNQSQG-PPPHGPKPERPPP-QGG-NOSHRPPPP-KPPER-PPQGQNQSOG-P-PP 186
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 QLCAQTARPPhASGPRRLGCERAWNHVSREAGVPLGLPAPGARRRGGGSARSRLPLPK 236
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 HPKGPEGPPQEGNKRSARSPPGKPQQGPPQEGNKKPQGGPKGPPPPGNGNPQQPQ 246
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 237 RPRGAAPERTVPVGSGWAHPGTRGPSDR-GFCVVSAPRAE-EATSLEGALSGSTRH 294
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 APPAGKPGQPPPPPGQGRPPRP 268
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 295 SHFSVGROHHAGPPPTSRRPP 316
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT      5
ID   FRP4_HUMAN STANDARD;          PRT;    247 AA.
AC   P10163; P02813;
DC   21-JUL-1986 (REL. 01, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DI   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   SALIVARY PROLINE-RICH PROTEIN PO PRECURSOR (ALLELE S).
GN   PRB4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
```

Dd 273 SVLPRTSPRRGLLNPNNRHRHTSTGHIPPTTTTSRTPGPPSKRLQPVH-LYQSSPHTNFRP 331
 : :: : : : : : : : : : : : : : : : :
 Qy 286 EG-ALSGTRHSHPSVGRQHAGPESTSRPRPWDTPCPPVYAETKFLY--SSGDKEQLRP 343
 :
 Dd 332 SSIRKDALLQTGPRLGHLERLGQPANLRTSERGPPPKRRLPRSEPNRLPKPLPEATLAP 391
 :
 Qy 344 SFLD-SSL-R--PSLTGARLVTETFGSLRPWMPGTFRRLPRLFPORYWQRPLFLELLG- 398
 :
 Dd 392 SYRHRRVPILLNPAPAALPSIATVSSRGKTHSLPKGALKPK-EGAPPPPRRL 442
 :
 Qy 399 NHAOC-PYGVLTKTHCPURA-AVTPAAGVCAREKPQGSVAPEEEDTDPRL 448
 :

```

RESULT      7
ID ID PRP2_MOUSE STANDARD;          PRT;    261 AA.
AC P05142;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROLINE-RICH PROTEIN MP-2 PRECURSOR.
GN PRP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 86059475.
RA ANN D.K., CARLSON D.M.;
J. BIOL. CHEM. 260:15863-15872(1985).
DR EMBL; M12099; G200547; -.
DR HSSP; P19999; ICLG.
REPEAT; SALIVA; SIGNAL.
FT SIGNAL 1 15
CHAIN 16 261
SQ SEQUENCE 261 AA; 26034 MW; 9D830DAF CRC32;
Query Match 1.5%; Score 125; DB 1; Length 261;
Best Local Similarity 27.0%; Pred. No. 3.14e-03;
Matches 40; Conservative 33; Mismatches 69; Indels 6; Gaps 5;

Dd 76 PQGGPPPGGQQPRPPQGPPPPGGGQQPRPPGGGQQPRPPGGGQQPRPPGGGQQPRPPQG 135
: : : : : : : : : : : : : : : : : : : :
Qy 181 ATQARPPPHASGPRRLGCERAMNHVSREAGVPLGLPAGARRRGGSSASRLPLKRRR 240
: : : : : : : : : : : : : : : : : : : :
Dd 136 PPPGGQLRPPQQPPPPAGP-QPRPPQGGPPPPAGPQPRPPGPTTGQPRPTQGPPT 194
: : : : : : : : : : : : : : : : : : : :
Qy 241 GAADPE-PE-RTPVQGGSWAHFGRTGRGSDRCFCVYSVARPAEEATSLEGALSCTRHSHP 298
: : : : : : : : : : : : : : : : : : : :
Dd 195 GGQQQQPPGPPPPGGGQQPRPPQGPPPP 222
: : : : : : : : : : : : : : : : : : : :
Qy 299 VGRGHA--GGPSTSRP-PRFWDTFCPP 323
: : : : : : : : : : : : : : : : : : : :
```

```

RESULT      8
ID ID PRP3_MOUSE STANDARD;          PRT;    296 AA.
AC P05143;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROLINE-RICH PROTEIN MP-3 (FRAGMENT).
GN PRP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 86059475.
RA ANN D.K., CARLSON D.M.;
J. BIOL. CHEM. 260:15863-15872(1985).
DR EMBL; M12100; G200549; -.
DR HSSP; P19999; ICLG.
REPEAT; SALIVA.
FT NON TER 1
KW NON TER 1
```


OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC	EUTHERIA; PRIMATES.
[1]	
RN	SEQUENCE FROM N.A.
RP	TISSUE=HIPPOCAMPUS;
RC	MEDLINE: 93094231.
RX	OEZKAYNAK E., SCHNEGELSBERG P.N.J., JIN D.F., CLIFFORD G.M.,
RA	WARREN F.D., DRIER E.A., OPPERMAN H.;
RA	J. BIOL. CHEM. 267:25220-25227(1992).
RL	-1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION. MAY BE THE
CC	OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF
CC	EPITHELIAL OSTEOGENESIS. PLAYS A ROLE IN CALCIUM REGULATION
CC	AND BONE HOMEOSTASIS (BY SIMILARITY).
CC	-1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC	-1- SIMILARITY: TO OTHER GROWTH FACTORS OF THE TGF-BETA FAMILY.
DR	EMBL; M57016; G189390; -.
DR	PIR; A45056; A45056.
DR	HSSP; P08112; ITPG.
DR	PROSITE; PS00250; TGF_BETA; 1.
KW	SIGNAL; GROWTH FACTOR; CYTOKINE; BONE; CARTILAGE; GLYCOPROTEIN.
FT	SIGNAL 1 19 POTENTIAL.
FT	PROPEP 20 263 POTENTIAL.
FT	CHAIN 264 402 BONE MORPHOGENETIC PROTEIN 8.
FT	DISULFID 301 367 BY SIMILARITY.
FT	DISULFID 330 399 BY SIMILARITY.
FT	DISULFID 334 401 BY SIMILARITY.
FT	DISULFID 366 366 INTERCHAIN (BY SIMILARITY).
FT	CARBOHYD 158 158 POTENTIAL.
FT	CARBOHYD 343 343 POTENTIAL.
SQ	SEQUENCE 402 AA; 44764 MW; 8F20C81B CRC32;
Query Match 1.5%; Score 125; DB 1; Length 402;	
Best Local Similarity 39.7%; Pred. No. 3.14e-03;	
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps	
Db	12 GLALCALGGGGLRPPP-GC-PQRRLG-ARERDVRQREILAVLGLPGR-PRRPPAPAA5
QY	172 GPPLYQLGAQTAA-RPPPHASGPRRLGCRAWNHSVREAGVPLGLPAPGARRRGGASR
Db	68 RLP 70
QY	231 SLP 233
RESULT 11	
ID	TS13 MOUSE STANDARD; PRT: 566 AA.
AC	Q01755;
DT	01-NOV-1995 (REL. 32, CREATED)
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DE	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE	TESTIS-SPECIFIC PROTEIN PBS13.
GN	TCPI1.
OS	MUS MUSCULUS (MOUSE).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC	EUTHERIA; RODENTIA.
[1]	
RP	SEQUENCE FROM N.A.
RC	STRAIN=CBA/CA; TISSUE=TESTIS;
RX	MEDLINE: 91372153.
RA	MAZARAKIS N.D., NELKI D., LYON M.F., EVANS E.P., RUDDY S.,
RA	FREEMONT P., DUDLEY K.;
RL	DEVELOPMENT 111:561-571(1991).
CC	-1- FUNCTION: POSSIBLY PLAYS AN IMPORTANT ROLE IN SPERM DEVELOPMENT
CC	AND FUNCTION.
CC	-1- TISSUE SPECIFICITY: TESTIS.
CC	-1- DEVELOPMENTAL STAGE: FIRSTLY EXPRESSED IN THE PACHYTENE
CC	SPERMATOCYTE STAGE.
CC	-1- SIMILARITY: SOME TO YEAST SOK1.
DR	EMBL; X52128; G54853; -.
DR	PIR; S22933; S22933.
DR	MGI; MGI:98544; TCPI1.
KW	TESTIS; SPERMATOGENESIS; REPEAT; COILED COIL.
FW	DOMAIN 171 206 COILED COIL (5 HEPTADS).

FT DOMAIN 256 318 COILED COIL (8 HEPTADS).
 FT DOMAIN 419 432 COILED COIL (2 HEPTADS).
 FT DOMAIN 440 509 COILED COIL (10 HEPTADS).
 FT DOMAIN 298 318 LEUCINE-RICH REPEATS.
 FT REPEAT 298 304 LRR 1.
 FT REPEAT 305 311 LRR 2.
 FT REPEAT 312 318 LRR 3.
 SQ SEQUENCE 566 AA; 61970 MW; 7A421F03 CRC32;

Query Match 1.5%; Score 127; DB 1; Length 566;
 Best Local Similarity 35.7%; Pred. No. 1.60e-03;
 Matches 30; Conservative 19; Mismatches 30; Indels 5; Gaps 5;

Db 34 SARGTDVSTVARARPPSPQGRRAVKTAAPRGVGHGLRTGPTSRCQPQSARAKLPS 93
 :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
 QY 217 PAPCARRRGGASRSLLP-PKRPGRGAPEPTVPVGSGS-WAHPP-GRTRGPSDRG-FCV 272
 :|::||::||::||::||::||::||::||::||::||::||::||::||::||:

Db 94 VTRCAGLPSPG-KHGLGTTPSSH 116
 :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
 Q 273 VSPARPAEATSLLEGALSGTRHS 296
 :|::||::||::||::||::||::||::||::||::||::||::||::||::||:

RESULT 12

ID MIAE_SALTY STANDARD; PRT; 270 AA.

AC Q08015; AC
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE TENA-(MS[2]TO[6]A)-HYDROXYLASE (EC 1.-.-.-).
 GN MIAE.
 OS SALMONELLA TYPHIMORIUM.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 CC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2.
 RX MEDLINE; 94075212.
 RA PERSON B.C.; BJOERK G.R.;
 RL J. BACTERIOLOG. 175:7776-7785(1993).
 CC -1- FUNCTION: PRODUCTION OF THE MODIFIED NUCLEOSIDE 2-METHYLTHIO-CIS-
 CC RIBOZEIN (MS[2]TO[6]A) FOUND IN SOME TNAS. CATALYZES THE
 CC OXYGEN-DEPENDENT TRANSFORMATION OF MS[2][6]A INTO MS[2]TO[6]A.
 CC -1- PATHWAY: THIRD STEP OF SYNTHESIS OF 2-METHYLTHIO-CIS-RIBOZEIN.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
 DR ENBL; X73368; G312708; --
 DR PIR; S34361; S34361.
 DR STYGENE; SG10234; MIAE.
 DR OXIDOREDUCTASE; TRNA PROCESSING.
 KW [REDACTED]
 SQ SEQUENCE 270 AA; 31140 MW; C528DF2 CRC32;

Query Match 1.4%; Score 116; DB 1; Length 270;
 Best Local Similarity 35.6%; Pred. No. 5.84e-02;
 Matches 16; Conservative 13; Mismatches 14; Indels 2; Gaps 2;

Db 15 LRMYNPQLSPVLNFLHCPTQAW-IYQARDPQNPLLLTLHLIC 58
 ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
 QY 14 LRSHYREVLP-LATFVRRLGPQGWRLVQRGDPAAFRALVAQCCLVC 57
 ||::||::||::||::||::||::||::||::||::||::||::||::||::||:

RESULT 13

ID Y091_NPVOP STANDARD; PRT; 279 AA.

AC Q10341;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 29.3 KD PROTEIN (ORF92).
 OS ORGYIA PSEUDOTSUGATA MULTICAPSID POLYHEDROSIS VIRUS (OPMNFV).
 CC VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ARENS C.A.; RUSSELL R.R.; FUNK C.J.; EVANS J.; HARWOOD S.,
 RA ROHRMANN G.F.;
 RL VIROLOGY 229:381-399(1997).

(TM)

Result No.	Score	Query Match	Query			ID	Description	Pred. No.
			Length	DB				
1	8465	100.0	1132	2	014746	TELOMERASE REVERSE TRA	0.00e+00	
2	8459	99.9	1132	2	014793	TELOMERASE CATALYTIC S	0.00e+00	
3	678	8.0	988	1	013339	TELOMERASE REVERSE TRA	6.18e-11	
4	681	8.0	989	1	013338	TELOMERASE REVERSE TRA	1.42e-11	
5	439	5.2	1031	3	000939	TELOMERASE SUBUNIT P12	3.85e-60	
6	424	5.0	67	10	035432	TELOMERASE CATALYTIC S	4.19e-57	
7	314	3.7	884	1	006133	CHROMOSOME XII COSMID	2.25e-35	
8	142	1.7	234	2	000600	PAROTID 'O' PROTEIN (F	3.44e-05	
9	141	1.7	276	9	046612	IS 1222 GENE ORF-A AND	4.83e-05	
10	147	1.7	296	11	069118	HYPOTHETICAL PROTEIN (6.21e-06	
11	143	1.7	574	13	036027	ACTIN ASSOCIATED PROTE	2.45e-05	
12	136	1.6	315	2	039076	HOMEOBOX PROTEIN (FRAG	2.58e-04	
13	132	1.6	317	10	Q82103	PROLINE RICH PROTEIN P	9.59e-04	
14	133	1.6	539	3	002133	SIMILARITY TO COLLAGEN	3.59e-04	
15	135	1.6	585	11	041935	HYPOTHETICAL 60.2 KD P	6.59e-04	
16	132	1.6	924	11	099307	LF3 PROTEIN.	9.59e-04	
17	125	1.5	227	10	Q62107	PROLINE-RICH SALIVARY	9.02e-03	
18	129	1.5	264	11	Q85301	HOMOLOGUE OF RETROVIRA	2.53e-03	
19	131	1.5	300	10	Q61888	PROLINE RICH PROTEIN.	1.33e-03	
20	129	1.5	309	2	Q04118	SALIVARY PROLINE-RICH	2.53e-03	

Db 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGKEQLRPSFLSSLRPSLTGARRL 360
QY 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGKEQLRPSFLSSLRPSLTGARRL 360
Db 361 VETIFGSRPMPGTPRLPRLPORYWQMPLELLEGNHAQCPYGVLLKTHCPRLRAVT 420
QY 361 VETIFGSRPMPGTPRLPRLPORYWQMPLELLEGNHAQCPYGVLLKTHCPRLRAVT 420
Db 421 PAAGVCAREKPOGSVAAPAEEDTDPRRLVOLLROHSSPWQVYGVRACLRLVPPGLWGS 480
QY 421 PAAGVCAREKPOGSVAAPAEEDTDPRRLVOLLROHSSPWQVYGVRACLRLVPPGLWGS 480
Db 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540
QY 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540
Db 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 600
QY 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 600
Db 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPIVNMDDYVVGARTFREKRAERLTSRVKA 660
QY 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPIVNMDDYVVGARTFREKRAERLTSRVKA 660
Db 661 LFSVLNTERARRPGLLGASVGLDLDIHRAWRTFVLRVRAQDPPPELYFYKVDVTGAYDTI 720
QY 661 LFSVLNTERARRPGLLGASVGLDLDIHRAWRTFVLRVRAQDPPPELYFYKVDVTGAYDTI 720
Db 721 PODRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRAFKSHVSTLDDLPYMRQFVAHL 780
QY 721 PODRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRAFKSHVSTLDDLPYMRQFVAHL 780
Db 781 QETSPLRDVAVVIEQSSSLEASGLFVFLRPMCHHVRIRKSVYOCQIGPOGSIILSTL 840
QY 781 QETSPLRDVAVVIEQSSSLEASGLFVFLRPMCHHVRIRKSVYOCQIGPOGSIILSTL 840
Db 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 1020
QY 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 1020
Db 1021 FHOQVKNPTFFLRVSDTASLCYSILKAKNAGSLGAKGAAGPLPSEAVQWLCHQAFLL 1080
QY 1021 FHOQVKNPTFFLRVSDTASLCYSILKAKNAGSLGAKGAAGPLPSEAVQWLCHQAFLL 1080
Db 1081 KLTRHRVYVPLLGSLRTAQTSRLKPLGTTLTALAAANPALPSDFKTILD 1132
QY 1081 KLTRHRVYVPLLGSLRTAQTSRLKPLGTTLTALAAANPALPSDFKTILD 1132

RESULT 2

ID 014783 PRELIMINARY; PRT: 1132 AA.
AC 014783;
DT 01-JAN-1998 (TREMBLEL. 05, CREATED)
DT 01-JAN-1998 (TREMBLEL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLEL. 05, LAST ANNOTATION UPDATE)
DE TELOMERASE CATALYTIC SUBUNIT.
GN HEST2.
OS HOMO SAPIENS (HUMAN).
OC EUMYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA MEYERSON M., COUNTER C.M., EATON E.N., ELLISEN L.W., STEINER P.,
RA CADDLE S.D., ZIAUGRA L., BEIJERSBERGEN R.L., DAVIDOFF M.J., LIU Q.,

RA BACCHETTI S., HABER D.A., WEINBERG R.A.;
RL CELL 90:785-795(1997).
DR EMBL; AF018167; G2347129; -.
SQ SEQUENCE 1132 AA; 126938 MW; C1E5E2AF CRC32;

Query Match 99.9%; Score 8459; DB 2; Length 1132;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MPRAPRCRAVRSLLRSHYREVLPPLATEFVRRLGPGWRLVQGDPAAPRALVAQCLVCVPW 60
QY 1 MPRAPRCRAVRSLLRSHYREVLPPLATEFVRRLGPGWRLVQGDPAAPRALVAQCLVCVPW 60
Db 61 DARPPPAAPSFQVSCLELVARVLQRLCERGAKNVLAFAFALLDARGGPEAFTSVR 120
QY 61 DARPPPAAPSFQVSCLELVARVLQRLCERGAKNVLAFAFALLDARGGPEAFTSVR 120
Db 121 SYLPTNTVDALRGSGAWGLLLRRVGGDDVLLHLLARCAFLVLAAPSCAYQVCGPPLYQLGA 180
QY 121 SYLPTNTVDALRGSGAWGLLLRRVGGDDVLLHLLARCAFLVLAAPSCAYQVCGPPLYQLGA 180
Db 181 ATOARPPPHASGPRRLGRCERAWNHSVREAGVPLGLPAGARRGGGSASRLPLPKPRR 240
QY 181 ATOARPPPHASGPRRLGRCERAWNHSVREAGVPLGLPAGARRGGGSASRLPLPKPRR 240
Db 241 GAAPPEPERTVPGQGSWAHPGRTGSPDRGFCVWSPARPAEATSLEGALSGTRHSHPSVG 300
QY 241 GAAPPEPERTVPGQGSWAHPGRTGSPDRGFCVWSPARPAEATSLEGALSGTRHSHPSVG 300
Db 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGKEQLRPSFLSSLRPSLTGARRL 360
QY 301 ROHAGPPSTSRPRPMDTPCPVYAEKHLFYSXSGKEQLRPSFLSSLRPSLTGARRL 360
Db 361 VETIFGSRPMPGTPRLPRLPORYWQMPLELLEGNHAQCPYGVLLKTHCPRLRAVT 420
QY 361 VETIFGSRPMPGTPRLPRLPORYWQMPLELLEGNHAQCPYGVLLKTHCPRLRAVT 420
Db 421 PAAGVCAREKPOGSVAAPAEEDTDPRRLVOLLROHSSPWQVYGVRACLRLVPPGLWGS 480
QY 421 PAAGVCAREKPOGSVAAPAEEDTDPRRLVOLLROHSSPWQVYGVRACLRLVPPGLWGS 480
Db 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540
QY 481 RHNERFLRNTKFFISLGKHAJLSQBELTWKMSVRCDAWLRRSPGVGCVPAAEHLRREEI 540
Db 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 600
QY 541 LAKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 600
Db 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPIVNMDDYVVGARTFREKRAERLTSRVKA 660
QY 601 LSAEVRQHREARPAALLTSRLRTPKPDGLRPIVNMDDYVVGARTFREKRAERLTSRVKA 660
Db 661 LFSVLNTERARRPGLLGASVGLDLDIHRAWRTFVLRVRAQDPPPELYFYKVDVTGAYDTI 720
QY 661 LFSVLNTERARRPGLLGASVGLDLDIHRAWRTFVLRVRAQDPPPELYFYKVDVTGAYDTI 720
Db 721 PODRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRAFKSHVSTLDDLPYMRQFVAHL 780
QY 721 PODRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRAFKSHVSTLDDLPYMRQFVAHL 780
Db 781 QETSPLRDVAVVIEQSSSLEASGLFVFLRPMCHHVRIRKSVYOCQIGPOGSIILSTL 840
QY 781 QETSPLRDVAVVIEQSSSLEASGLFVFLRPMCHHVRIRKSVYOCQIGPOGSIILSTL 840
Db 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 900
QY 841 LKFLHLMSSVYVVELLSRFFYTTTQKNRUFFYRKSVWSKLQSIGIROHLKRVOLRE 900
Db 901 RKTVVNFPVEDEALGGTAFAVQMPAHGLFPWCGLLLDTRTLEVOGSDYSYARTSTRASLT 960
QY 901 RKTVVNFPVEDEALGGTAFAVQMPAHGLFPWCGLLLDTRTLEVOGSDYSYARTSTRASLT 960

Db 961 NRGFKAGNRMRKLVGLRKLKCHSLFDLOVNSLOTVCTNYIKILLLOAYRFHACVLOLP 1020
 QY 961 NRGFKAGNRMRKLVGLRKLKCHSLFDLOVNSLOTVCTNYIKILLLOAYRFHACVLOLP 1020
 Db 1021 FHOQVKNFTFFLRVSDTASLCYSILKAKNAGMSLGAKAGPLPSEAYOWLCHQAFLL 1080
 QY 1021 FHOQVKNFTFFLRVSDTASLCYSILKAKNAGMSLGAKAGPLPSEAYOWLCHQAFLL 1080
 Db 1081 KLTRHRTVYVPLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSPDKTILD 1132
 QY 1081 KLTRHRTVYVPLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSPDKTILD 1132
 RESULT 3
 ID O13339 PRELIMINARY; PRT; 988 AA.
 AC O13339;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE TELOMERASE REVERSE TRANSCRIPTASE 1.
 GN TRT1.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
 RA LINGNER J., HARLEY C.B., CECI T.R.;
 RL SCIENCE 277:955-959(1997).
 DR EMBL; AF015783; G2340168; -.
 KW RNA-DIRECTED DNA POLYMERASE.
 SQ SEQUENCE 988 AA; 116328 MW; 39C385A7 CRC32;

Query Match 8.0%; Score 678; DB 1; Length 988;
 Best Local Similarity 26.4%; Pred. No. 6.18e-110;
 Matches 129; Conservative 133; Mismatches 208; Indels 18; Gaps 15;

Db 340 PNOVFAFLRSILVRVFPKLTWGNQRIEILKOLETFLKLSRYESFSLHYLMSNIKISEI 399
 QY 458 PQOYGVFVRACLRLVPPGLWGRSHNRERFLRNTKFKISLGHAKLSQELTWMKSVRDC 517
 Db 400 EWLVLGRSNAMKCLSDFEKR-KQ-IPAEFYIWLNSFIPILOSFYFITESDLNRNTV 457
 QY 518 AWL---RRSPGVGCVPAAEHLREELAKFLHLMVSVVVELLSRFFVTETTFQKNRLF 574
 Db 458 YFRKDIW-KLLCRPFITSMKMEAFKINENNVRMDTQ-KITLPPAVIRLLPKKNTFRLIT 515
 QY 575 FYKSVWSKLQSIGIRHQLKRVQLRELSEAEVROHREARPALTSRLRFPKDGRLPIV 634
 Db 516 NLKRFELIKMGSKNK-LMVLSTNOTLRPVASILKHLNEESSGIPFN-LEVYMKLLTFKKDL 574
 QY 635 NMDYVVGATFREKRAERLTSVKALFSLVLYERARRPGLGASVGLDDIHRWRTFV 694
 Db 575 LKRMFGRK-K-YFVRIDIKSCYDRIDQDLFRIVKKKLKDPE-FVIRKYATIH-ATS 630
 QY 695 LRVRADQPPPELYFVKVDVTGAYDTIPQDRLEVIASIIKPTQYTCVRRYAVVQAAHGH 754
 Db 631 ATKNFVSEAFSFDVMPFEK--VVOLLSMKT--SDTLFVDVFDYTWKSSSEIFKMLKEHLS 687
 QY 755 VRKAFKSHVSTLTDLPQYMRQFVAHLQETSPLRDVAVIEQSSSLNEASSGLFDVFLRMC 814
 Db 688 GHIVKIGNSQYLQKVGIPQGSILSFLCHFYMEDLIDEVLSFTKKKGVLLRVVDDFLFI 747
 QY 815 HIAVRIRGKSVQCOGIPQGSILSFLCHFYMEDLIDEVLSFTKKKGVLLRVVDDFLFI 873
 Db 748 TVNKKDAKFLNLSRGFEKHNFTSLEKTVINFNENSGIINNTFFNESKKR-M-PFFGF 805
 QY 874 TPLHTHAKTFLRLTVRGVPEYGCVVNLKRTVNVFPVEDEALGCTAFVQMPAHGLFPWGL 933
 Db 806 SVNMRSLD 813
 QY 934 LLDTRTLE 941

RESULT 4
 ID O13338 PRELIMINARY; PRT; 989 AA.
 AC O13338;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE TELOMERASE REVERSE TRANSCRIPTASE 1.
 GN TRT1.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
 RA LINGNER J., HARLEY C.B., CECI T.R.;
 RL SCIENCE 277:955-959(1997).
 DR EMBL; AF015783; G2340169; -.
 KW RNA-DIRECTED DNA POLYMERASE.
 SQ SEQUENCE 989 AA; 116456 MW; FDE74202 CRC32;

Query Match 8.0%; Score 681; DB 1; Length 989;
 Best Local Similarity 26.3%; Pred. No. 1.42e-110;
 Matches 129; Conservative 136; Mismatches 204; Indels 21; Gaps 18;

Db 340 PNOVFAFLRSILVRVFPKLTWGNQRIEILKOLETFLKLSRYESFSLHYLMSNIKISEI 399
 QY 458 PQOYGVFVRACLRLVPPGLWGRSHNRERFLRNTKFKISLGHAKLSQELTWMKSVRDC 517
 Db 400 EWLVLGRSNAMKCLSDFEKR-KQ-IPAEFYIWLNSFIPILOSFYFITESDLNRNTV 457
 QY 518 AWL---RRSPGVGCVPAAEHLREELAKFLHLMVSVVVELLSRFFVTETTFQKNRLF 574
 Db 458 YFRKDIW-KLLCRPFITSMKMEAFKINENNVRMDTQ-KITLPPAVIRLLPKKNTFRLIT 515
 QY 575 FYKSVWSKLQSIGIRHQLKRVQLRELSEAEVROHREARPALTSRLRFPKDGRLPIV 634
 Db 516 NLKRFELIKMGSKNK-LMVLSTNOTLRPVASILKHLNEESSGIPFN-LEVYMKLLTFKK 573
 QY 635 NMDYVVGATFREKRAERLTSVKALFSLVLYERARRPGLGASVGLDDIHRWRTFV 692
 Db 574 LKRMFGRK-K-YFVRIDIKSCYDRIDQDLFRIVKKKLKDPE-FVIRKYATIH-ATS 629
 QY 693 FLVRADQPPPELYFVKVDVTGAYDTIPQDRLEVIASIIKPTQYTCVRRYAVVQAAH 752
 Db 630 DRATKNFVSEAFSFDVMPFEK--VVOLLSMKT--SDTLFVDVFDYTWKSSSEIFKMLKEH 686
 QY 753 GHVRKAFKSHVSTLTDLPQYMRQFVAHLQETSPLRDVAVIEQSSSLNEASSGLFDVFLRF 812
 Db 687 LSGHIVKIGNSQYLQKVGIPQGSILSFLCHFYMEDLIDEVLSFTKKKGVLLRVVDDFL 746
 QY 813 MCHAVIRGKSVQCOGIPQGSILSFLCHFYMEDLIDEVLSFTKKKGVLLRVVDDFL 871
 Db 747 FITVNNKDAKFLNLSRGFEKHNFTSLEKTVINFNENSGIINNTFFNESKKR-M-PFF 804
 QY 872 LVTPHHTHAKTFLRLTVRGVPEYGCVVNLKRTVNVFPVEDEALGCTAFVQMPAHGLFPWC 931
 Db 805 GFSVNMRSLD 814
 QY 932 GLLDTRTLE 941

RESULT 5
 ID O00939 PRELIMINARY; PRT; 1031 AA.
 AC O00939;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE TELOMERASE SUBUNIT P123.
 OS EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; ALVEOLATA; CILIOPHORA;
 OC EUPLOTES; EUPLOTIDA; EUPLOTES.
 RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE; 97274210.
RA LINGNER J., HUGHES T.R., SHEVCHENKO A., MANN M., LUNDBLAD V.,
RA CECH T.R.;
RL SCIENCE 276:561-567(1997).
DR EMBL; U95964; G2072336; -.
SQ SEQUENCE 1031 AA; 122562 MW; 21A885CD CRC32;

Query Match
Best Local Similarity 5.2%; Score 439; DB 3; Length 1031;
Best Local Similarity 23.0%; Pred. No. 3.85e-60;
Matches 142; Conservative 168; Mismatches 261; Indels 47; Gaps 34;

Db 361 FINEFFNIILPKDLTGR-NRKNFQKKVKKYVELNKHIE-LIHKNLLEKINTREISIMQV 418
QY 464 FVACRLRLVPPGLGSRHNRERLRNKKFISLGKAKLSQELTW-KMSVRDCAMLR 522
Db 419 ETSAKHEFFYFDHE-NIVVLKLLRWIFEDLVSLIRFFVYVEQOKSVSKTYIRKNLWD 477
QY 523 SPGVGCVPAEHRLEBILAKFLHWSVYVELRFFVYTFQKNRUFFYRKSVWS 582
478 VIMKMSI-ADLKRETLAEVQEKY-EEMKSLGAPAGKRLRIPKTTFRPI--MTF--NK 531
QY 583 KLASIGIRQHLKRVQLSELSEAEVQHRARPALLTSRLRIPKDPGLRPVNDYVVG 642
Db 532 KYNSDRKTKLTNTKLLSHMLKTLKNMFKDPGFVFNVDVDMKKYEEFVCK-WK 590
QY 643 RTRFRKRAERLTSRVKALFS--VINYERARR-PGLLGASVLGLDDIHRAWRTFVLVRA 699
Db 591 QVGOPKLEFFATMDIEKCYDSVNRKELSTFKTKLLSSDFWIMTAQILKRNKNIVIDSKN 650
QY 700 QDPPPELYFVKVDVTGAYDIPQRLTEVI-AS-IIPQ---NYCV--RYAVV-Q-KA 750
Db 651 FRKEMKDYFRQKFKIAEGGYPTFLSVLENQDNLNAKKTLLIVEAKQRYFKKNL 710
QY 751 AHGHVKA-FKSHVSTLT-DLOPYMQF-V-AHQETSPLRDVAVIEQSSSLNEASSGLF 806
Db 711 QPVIN-ICQVYNINFGKFKYQTKIGIQGLCVSSILSFFYATLEESLGLFRLDESMP 769
QY 807 DVLFRPMCHHA-VRIKRSQVQCGIQPGISILSTLLGSLCYGDMENK-L-F--AGIRRD 860
Db 770 NPVNLMLRTDDYLLITTOENNAVLFIKILNVSRENGKFNKMKLOTSPPLSPSFAK 829
QY 861 G----LLRLVDDFLVTPHLLTHAKTLRLTVRGVPEYGVVNLKTVNPPVEDEALG- 915
Db 830 YGMDSVEEQNIQVDCWDIGISIDMKTLALPNIN-LRIEGILCTNLNMQTKASMWLK 888
QY 916 -GTAFFVQMPAHL-F-PWCGLLDTRILEVQSDYSSYARTSIRASLTFN-RGFRAGNMR 971
889 KLUKSLF-MNNITHYFKTTTTFDANKTLNKLFIISGYKYMCAKE--YKDHFKKNLAM 945
QY 972 RKLFGVRLKCHSLFLDLQVNSLQTVCTNIVYKILLQAYRFHACVLQLPFHQVKNPTE 1031
Db 946 SSMIDLEVSIIIVSUTRA 963
QY 1032 FLRVISDTASLCYSILKA 1049

RESULT 6
ID O35432 PRELIMINARY; PRT; 67 AA.
AC O35432;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE TELOMERASE CATALYTIC SUBUNIT (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
RN [1]
RP SEQUENCE FROM N.A.
RA DRISS R., CLEVELAND J.L.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF029235; G2605903; -.
FT NON_TER 1 1

ET NON_TER 67 67
SQ SEQUENCE 67 AA; 8368 MW; E2A06F2B CRC32;

Query Match
Best Local Similarity 5.0%; Score 424; DB 10; Length 67;
Matches 51; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Db 1 FFYVTESTFQKNRUFFYRKSVWSKLSQISGVQHLRLVRLRSLQSEVVRHODTWLAMPIC 60
QY 560 FFYVTESTFQKNRUFFYRKSVWSKLSQISGVQHLRLVRLRSLQSEVVRHODTWLAMPIC 619
Db 61 RLREIPK 67
QY 620 RLREIPK 626

RESULT 7
ID Q06163 PRELIMINARY; PRT; 884 AA.
AC Q06163;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE CHROMOSOME XII COSMID 8543.
GN L8543.12.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA DU Z.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FAVELLO A., FULTON L., GAITUNG S., GRECO T., KIRSTEN J.,
RA KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,
RA JOHNSON D., JOHNSON L., LANGSTON Y., LATREILLE P., LE T.,
RA MARDIS E., MENZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
RA RIFKEN L., RIFKEN L., TAICH A., TREVASKIS E., VIGNATI D.,
RA WILCOX L., WOULDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA WATERSTON R.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U20618; G662136; -.
SQ SEQUENCE 884 AA; 102663 MW; 1A94320F CRC32;

Query Match
Best Local Similarity 3.7%; Score 314; DB 1; Length 884;
Matches 118; Conservative 122; Mismatches 202; Indels 44; Gaps 37;

Db 271 LSHLSRQSPKER-VLKFTIIVILQKLPOEMFGSKKKKGIKKNLNLSPNGLYLPDS 329
QY 448 LVQLLRQHSSPWQVYGVFRACLRRLVPPGLGSRHNRERLNTKFTSLGKHAKLSQL 507
Db 330 LLKKLRKDFRWLFIS-DIWFTKHNFNQLAIC-FISWLFROLIPIOTFFYCTEIS 387
QY 508 LTWKMSVRDCAMLRSPGVCVPAEHRLEBILAKFLHWSVYVELRFFVYTFETT 567
Db 388 STVTIVVF-RHDTWNKLTIPFIVEYKTY-LVE-NNV-CRNHNSYTLNENHSMRIIPK 443
QY 568 FQKNRUFFYRKSVWSKLSQISGVQHLRLVRLRSLQSEVVRHODTWLAMPIC 626
Db 444 KSNNEFRITAIIPCRGADEEFTIYKENHK-NA-IQPTQKILEYLRNKRPTSF-TKIYSPT 500
QY 627 -PDGLRPVNDY-VVGARTFRREKRAERLTSRVKALFSVLNRYERARRPGLLGASVLGLD 684
Db 501 QIADRI-KEFKQRLKLLKNNVLPPELYFMKFDVSKYSDIPRMCMR-ILKDALKNENGFF 558
QY 685 DI-HRAWRTFVLV-RA-QDPPPELYFVKVDVTGAYDITPQ-DRLTEVIAIISIKPQNTYC 740

Db 559 VRQY-FFNTNT-G-VLKLENV-VNASRKPXY-ELYIDNVR-TVHLSNQDVIN---VV- 608
 QY 741 VR-RYAVVQRAAHGHVKAQKSHVSTITLDQYMRQFVAHLQETSPRLDAVVIEQSSLN 799
 Db 609 E---M-EIF-K-T---ALWEDKCYIREDFLQSGSSLPVLDVDDLLLEFYSEKAS 658
 QY 800 EASGGLDFVLRFWCHAVIRGKSYVQCQIPGSGILSTLLCSLCYQDM-E-NKLP-AG 856
 Db 659 PSQTLIKLADDFLIISTDQOQVINIKKLAMGQFQRYNAKAN-RDKILAVSSQSDDDTV 717
 QY 857 IRDGLLLRLVDVDFLLVTPHLTHAKTELRLTVRGVPEVGVNLRKTVVNFVPEDEALGG 916
 Db 718 IQFCAM 723
 QY 917 TAFVQM 922

RESULT 8
 ID Q00600 PRELIMINARY; PRT; 234 AA.
 Q00600;
 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
 DT PAROTID 'O' PROTEIN (FRAGMENT).
 GN PRB4.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RX MEDLINE; 96108975.
 RA AZEN E.A., AMBERGER E., FISHER S., PRAKOBPHOL A., NIECE R.L.;
 RL AM. J. HUM. GENET. 58:143-153(1996).
 DR EMBL; S80916; G1911492; -.
 FT NON_TER 1
 SQ SEQUENCE 234 AA; 23656 MW; B5920075 CRC32;

Query Match 1.7%; Score 142; DB 2; Length 234;
 Best Local Similarity 28.2%; Pred. No. 3.44e-05;
 Matches 40; Conservative 41; Mismatches 52; Indels 9; Gaps 9;
 Db 92 QGQNQSQG-PPPHGKPERPPP-OGG-NQSHRPPPPP-GKPER-PPFQGGNQSQG-P-PP 144
 QY 177 QLGAATQARPPPHGASGRRRLGRCERAWNHSVREAGVPLGLPAPGARRRGGSASRLPLK 236
 Db 145 HPKPEGPQQEGNKSARSPPKPGQPGQEQGNKPGQPPPPKPGQPPPPGPNQPOQ 204
 QY 237 RPRGAPEPERTPVGQSWAHPGTRGPSDR-GFCVVSAPPAE-EATSLGALSQTRH 294
 Db 205 APPAGKPGPPPPPGGGRPRP 226
 QY 295 SHPSVGRQHAGHPSTSRPRP 316

RESULT 9
 ID Q46612 PRELIMINARY; PRT; 276 AA.
 Q46612;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
 DE IS 1222 GENE ORF-A AND ORF-B.
 OS ENTEROBACTER AGGLOMERANS.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KLEBERGER, 1983;
 RX MEDLINE; 95255664.
 RA STEIBL H.D., LEWECKE F.M.;
 RL GENE 156:37-42(1995).
 RN [2]
 RP SEQUENCE FROM N.A.

RA STEIBL H.D., SIDDAVATTAM D.;
 RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA STEIBL H.D., SIDDAVATTAM D., KLINGMUELLER W.;
 RL PLASMID 34:223-228(1995).
 RN [4]
 RP SEQUENCE FROM N.A.

RA STEIBL H.D.;
 RL THESIS (1995), UNIVERSITAET BAYREUTH.
 DR EMBL; X78052; G459248; -.
 DR EMBL; X81893; E258949; -.
 SQ SEQUENCE 276 AA; 31718 MW; 0B54A420 CRC32;

Query Match 1.7%; Score 141; DB 9; Length 276;
 Best Local Similarity 33.7%; Pred. No. 4.83e-05;
 Matches 30; Conservative 23; Mismatches 30; Indels 6; Gaps 6;

Db 42 ITELALERRR-FGYRR-IWQLLRREGLHVNKRVYRLVHLSGLGVKRRRR-RKGLATERL 98
 QY 563 VTETFOKNRLLFFYRKSVWSKLSIGIGIRHLKRV-QLRELSEAEVQRHREARPAALLTSRL 621
 Db 99 PLL-RPAAPNLTSWSDDFVMDALATGRRIK 126
 QY 622 RFIPKPDGLRPVNMVYVGA-RTFRREK 649

RESULT 10
 ID Q69118 PRELIMINARY; PRT; 296 AA.
 Q69118;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN (FRAGMENT).
 OS HUMAN HERPESVIRUS TYPE 4.
 OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE;
 OC GAMMAHERPESVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87284169.
 RA FITZNER A.J., TSAI E.C., STROMINGER J.L., SPECK S.H.;
 RL J. VIROL. 61:2902-2909(1987).
 DR EMBL; M17294; G807646; -.
 KW HYPOTHETICAL PROTEIN.
 FT NON_TER 1
 SQ SEQUENCE 296 AA; 31393 MW; D17CEF6F CRC32;

Query Match 1.7%; Score 147; DB 11; Length 296;
 Best Local Similarity 38.6%; Pred. No. 6.21e-06;
 Matches 39; Conservative 20; Mismatches 33; Indels 9; Gaps 8;

Db 91 GPPPTRSGAAQ-RTHRRPPGCRSARNPGCPRTWRR--R-SGAQRHPPPGAGQRPSP 146
 QY 172 GPPLYQLGAATQARPPPHASG-PRR-R-LGGERAWNHSVREAGVPLGLPAPGA-RRRGGS 227
 Db 147 TGRPAAPGAPTAAAPGCGGAAPVSGATPHPERGSPAD 187
 QY 228 ASRSLPLKPRRGAAPER-TPVQGSWAHPGTRGPSD 267

RESULT 11
 ID Q36027 PRELIMINARY; PRT; 574 AA.
 Q36027;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE ACTIN ASSOCIATED PROTEIN.
 GN SPAC4F10.15C.
 OS UNKNOWN.
 OC UNCLASSIFIED.
 RN [1]
 RP SEQUENCE FROM N.A.
 RP CONNOR R., CHURCHER C.M.;

RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]
RA BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RC SEQUENCE FROM N.A.
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: 298980; E339884;
SQ SEQUENCE 574 AA; 59605 MW; 9E84D37C CRC32;

Query Match 1.7%; Score 143; DB 13; Length 574;
Best Local Similarity 28.8%; Pred. No. 2.45e-05;
Matches 42; Conservative 41; Mismatches 53; Indels 10; Gaps 7;

Db 305 AANKRRPPPP-PSRR---NRG-KPIGNGSSNSLPPPPPSRNASAG-SIDLPPQGR 358

QY 180 AATQARPPPHASGRRRLGCE-RAWNHSVREAGVPLGLPAPGARRRGSASRSLLPKRPR 239

Db 359 SAPPPPPPSAPSGROPPPLSSSRVSNPP--APPAPGSRAPALPLGNASRTSTPP 416

QY 240 RGAPEPRTVPGGSAWHP--GRTGFSDRGFCVVSAPAEATSLGALSGTNRHSHPS 298

417 VPTPP-SLPPSAPPSLPPSPAPSLPM 441

QY 299 VGRQHAGPSTSRPPRWDTPCPVP 324

RESULT 12
ID Q99076 PRELIMINARY; PRT; 316 AA.

AC Q99076; DB 13; Length 316; Score 132; DB 10; Length 317;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE HOMEBOX PROTEIN (FRAGMENT).

GN HB9.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-TONSIL;

RX MEDLINE: 91305125.

RA DEGUCHI Y., KEHL J.H.;

RL NUCLEIC ACIDS RES. 19:3742-3742(1991).

CC -I- SIMILARITY: STRONG WITH OTHER "ANTP-TYPE" HOMEBOX PROTEINS.

DR EMBL: X56537; E30256; ALT_SEQ.

DR PIR: S16681; S16681.

KW DNA-BINDING; NUCLEAR PROTEIN; HOMEBOX; PHOSPHORYLATION.

FT NON_TER 1

DNA_BIND 70 129 HOMEBOX.

DNA_BIND 112 121 H-T-H MOTIF.

SEQUENCE 316 AA; 34713 MW; 1531E766 CRC32;

Query Match 1.6%; Score 136; DB 2; Length 316;

Best Local Similarity 30.9%; Pred. No. 2.58e-04;

Matches 21; Conservative 20; Mismatches 24; Indels 3; Gaps 3;

QY 202 RPNSTASSDGVGTGTAAGPAT--SPRPSRPAQDRSPARSAPGPAAGPGGAWTH 260

QY 201 RANHSVREAGV-PLGLPAPGARRRGSASRSLLPKRPRG-AAPEPRTVGGSAW 258

Db 261 PAREPQA 268

QY 259 PGRTRGPS 266

RESULT 13
ID Q62103 PRELIMINARY; PRT; 317 AA.

AC Q62103; DB 2; Length 317; Score 133; DB 3; Length 539;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE PROLINE RICH PROTEIN PRECURSOR.

GN FRP.

OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CD-1; TISSUE-LIVER;

RX MEDLINE: 88273214.

RA ANN D.K., SMITH M.K., CARLSON D.M.;

RL J. BIOL. CHEM. 263:10867-10893(1988).

DR EMBL: M23236; G567232; -

DR MGD; MGI:97773; PRP.

FT SIGNAL. 1 15 POTENTIAL.

FT CHAIN 16 317 PROLINE-RICH PROTEIN.

SO SEQUENCE 317 AA; 31719 MW; 10C84341 CRC32;

Query Match 1.6%; Score 132; DB 10; Length 317;

Best Local Similarity 26.4%; Pred. No. 9.59e-04;

Matches 39; Conservative 38; Mismatches 54; Indels 7; Gaps 6;

Db 37 SGSPRRPVNSQGGPPPGQPRP-PQGGPPPGQPRPPQGGPPPGQPRP-PQGP 94

QY 180 AATQARPPPHAS--GPRRLGCE-RAWNHSVREAGVPLGLPAPGARRRGSASRSLLPK 236

Db 95 PPGGQPRPPQGGPPPGG-PQPRPPQGGPPPGQPRPPQGGPPPGQPRPPQGGPP 153

QY 237 RPRGAAPERTVPGGSAWHP--GRTGFSDRGFCVVSAPAEATSLGALSGTNRH 295

Db 154 GGQPRPPQGGPPPGQPRPPQGGPPPGP 181

QY 296 HPSVGRQHAGPSTSRPPRWDTPCPVP 323

RESULT 14

ID O02123 PRELIMINARY; PRT; 539 AA.

AC O02123; DB 2; Length 539; Score 133; DB 3; Length 539;

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)

DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

DT 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)

DE SIMILARITY TO COLLAGENS.

GN W03D2.1.

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,

RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,

RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,

RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,

RA LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B.,

RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFKEN L., ROOPRA A.,

RA SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E.,

RA STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,

RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,

RA WILKINSON-SPOAT J., WOHLDMAN P.;

RL NATURE 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA ROHLFING T., WOHLDMANN P.;

RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA WATERSTON R.;

RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AF000298; G1947160; -

SQ SEQUENCE 539 AA; 52555 MW; 78FF9A09 CRC32;

Query Match 1.6%; Score 133; DB 3; Length 539;
Best Local Similarity 31.7%; Pred. No. 6.92e-04;

[illegible]

RESULT	15
ID	O41935
AC	PRELIMINARY; PRT: 585 AA.
DT	O41935;
DT	01-JAN-1998 (TREMBREL. 05, CREATED)
DT	01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DT	01-JAN-1998 (TREMBREL. 05, LAST ANNOTATION UPDATE)
GN	HYPOTHETICAL 60.2 KD PROTEIN. GAMMAHV.M6.
OC	MURINE HERPESVIRUS 68.
OS	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-WUMS;
RX	MEDLINE; 97366649.
RA	VIRGIN H.W. IV, LATREILLE P., WAMSLEY P., HALLSWORTH K., WECK K.E., DAL CANTO A.J., SPECK S.H.;
RA	J. VIROL. 71:5894-5904(1997).
RL	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN-WUMS;
RC	LATREILLE P., WAMSLEY P., WATERSTON R.H.;
RA	SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR	EMBL; U97553; G2317934; -
KW	HYPOTHETICAL PROTEIN.
SQ	SEQUENCE 585 AA; 60160 MW; B35C73E5 CRC32;

```

Query Match      1.6%; Score 135; DB 11; Length 585;
Best Local Similarity 27.6%; Pred. No. 3,59e-04;
Matches 45; Conservative 37; Mismatches 68; Indels 13; Gaps 11;

Db      342 PBP-ELGCGSPSPAPSRAGARIPDLPGPLPSWGPPDRPPELPGCGSPSPAPSRAG 400
QY      173 PPIYUQGAATQARPPPHASGPRRR-L-GGERAWNHSVREA-GVP-LGLPAPARRRGUSA 228
      : : : : : : : : : : : : : : : : : : : : : : : : : :
401 ARIPDLGGLPSWGPPDRPPELPGP-GSPST-SPAPSRAGARIPRPSRPPPELGP--G 456
      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      229 SRSLPLPKR-PRGAAPAEPIRPVGGGSWAHFGRTGSPDSRCFCVVSAPRAEATISLEG 287

Db      457 SPFRPPRSRPP--PELPGCGSPDLGGLPSWG-PDPPTFPAPER 496
QY      288 ALSGTRHSHPISVGRQHHAHPGSTSRSPRPWDTPCPPIYAAETKH 430

```

Search completed: Thu Jul 30 14:51:34 1998
Job time : 147 secs.